

**Modélisation du dossier patient par deep learning : application à un  
entrepôt de données cliniques français pédiatrique et de maladies rares**

# **T H E S E**

**Présentée et publiquement soutenue devant**

**LA FACULTÉ DES SCIENCES MEDICALES ET PARAMEDICALES  
DE MARSEILLE**

**Le 26 Janvier 2022**

**Par Madame Mathilde BESSON**

**Née le 22 juin 1989 à Les Lilas (93)**

**Pour obtenir le grade de Docteur en Médecine**

**D.E.S. de SANTÉ PUBLIQUE ET MÉDECINE SOCIALE**

**Membres du Jury de la Thèse :**

<b>Monsieur le Professeur BOYER Laurent</b>	<b>Président</b>
<b>Monsieur le Professeur CHAMPSAUR Pierre</b>	<b>Assesseur</b>
<b>Madame le Professeur BERBIS Julie</b>	<b>Assesseur</b>
<b>Monsieur le Docteur ANTONINI François</b>	<b>Assesseur</b>
<b>Monsieur le Docteur NEURAZ Antoine</b>	<b>Directeur</b>



**Modélisation du dossier patient par deep learning : application à un  
entrepôt de données cliniques français pédiatrique et de maladies rares**

# **T H E S E**

**Présentée et publiquement soutenue devant**

**LA FACULTÉ DES SCIENCES MEDICALES ET PARAMEDICALES  
DE MARSEILLE**

**Le 26 Janvier 2022**

**Par Madame Mathilde BESSON**

**Née le 22 juin 1989 à Les Lilas (93)**

**Pour obtenir le grade de Docteur en Médecine**

**D.E.S. de SANTÉ PUBLIQUE ET MÉDECINE SOCIALE**

**Membres du Jury de la Thèse :**

<b>Monsieur le Professeur BOYER Laurent</b>	<b>Président</b>
<b>Monsieur le Professeur CHAMPSAUR Pierre</b>	<b>Assesseur</b>
<b>Madame le Professeur BERBIS Julie</b>	<b>Assesseur</b>
<b>Monsieur le Docteur ANTONINI François</b>	<b>Assesseur</b>
<b>Monsieur le Docteur NEURAZ Antoine</b>	<b>Directeur</b>

## FACULTÉ DES SCIENCES MÉDICALES & PARAMÉDICALES

<b>Doyen</b>	:	<b>Pr. Georges LEONETTI</b>
Vice-Doyen aux affaires générales	:	Pr. Patrick DESSI
Vice-Doyen aux professions paramédicales	:	Pr. Philippe BERBIS
Conseiller	:	Pr. Patrick VILLANI

### **Asseseurs :**

➤ aux études	:	Pr. Kathia CHAUMOITRE
➤ à la recherche	:	Pr. Jean-Louis MEGE
➤ à l'unité mixte de formation continue en santé	:	Pr. Justin MICHEL
➤ pour le secteur NORD	:	Pr. Stéphane BERDAH
➤ Groupements Hospitaliers de territoire	:	Pr. Jean-Noël ARGENSON
➤ aux masters	:	Pr. Pascal ADALIAN

### **Chargés de mission :**

➤ sciences humaines et sociales	:	Pr. Pierre LE COZ
➤ relations internationales	:	Pr. Stéphane RANQUE
➤ DU/DIU	:	Pr. Véronique VITTON
➤ DPC, disciplines médicales & biologiques	:	Pr. Frédéric CASTINETTI
➤ DPC, disciplines chirurgicales	:	Dr. Thomas GRAILLON

## ÉCOLE DE MEDECINE

<b>Directeur</b>	:	<b>Pr. Jean-Michel VITON</b>
------------------	---	------------------------------

### **Chargés de mission**

▪ PACES – Post-PACES	:	Pr. Régis GUIEU
▪ DFGSM	:	Pr. Anne-Laure PELISSIER
▪ DFASM	:	Pr. Marie-Aleth RICHARD
▪ DFASM	:	Pr. Marc BARTHET
▪ Préparation aux ECN	:	Dr Aurélie DAUMAS
▪ DES spécialités	:	Pr. Pierre-Edouard FOURNIER
▪ DES stages hospitaliers	:	Pr. Benjamin BLONDEL
▪ DES MG	:	Pr. Christophe BARTOLI
▪ Démographie médicale	:	Dr. Noémie RESSEGUIER
▪ Etudiant	:	Elise DOMINJON

## ÉCOLE DE DE MAIEUTIQUE

**Directrice** : **Madame Carole ZAKARIAN**

**Chargés de mission**

- 1<sup>er</sup> cycle : Madame Estelle BOISSIER
- 2<sup>ème</sup> cycle : Madame Cécile NINA

## ÉCOLE DES SCIENCES DE LA RÉADAPTATION

**Directeur** : **Monsieur Philippe SAUVAGEON**

**Chargés de mission**

- Masso- kinésithérapie 1<sup>er</sup> cycle : Madame Béatrice CAORS
- Masso-kinésithérapie 2<sup>ème</sup> cycle : Madame Joannie HENRY
- Mutualisation des enseignements : Madame Géraldine DEPRES

## ÉCOLE DES SCIENCES INFIRMIERES

**Directeur** : **Monsieur Sébastien COLSON**

**Chargés de mission**

- Chargée de mission : Madame Sandrine MAYEN RODRIGUES
- Chargé de mission : Monsieur Christophe ROMAN

## PROFESSEURS HONORAIRES

MM	AGOSTINI Serge	MM	DEVIN Robert
	ALDIGHERI René		DEVRED Philippe
	ALESSANDRINI Pierre		DJIANE Pierre
	ALLIEZ Bernard		DONNET Vincent
	AQUARON Robert		DUCASSOU Jacques
	ARGEME Maxime		DUFOUR Michel
	ASSADOURIAN Robert		DUMON Henri
	AUFFRAY Jean-Pierre		ENJALBERT Alain
	AUTILLO-TOUATI Amapola		FAUGERE Gérard
	AZORIN Jean-Michel		FAVRE Roger
	BAILLE Yves		FIECHI Marius
	BARDOT Jacques		FARNARIER Georges
	BARDOT André		FIGARELLA Jacques
	BERARD Pierre		FONTES Michel
	BERGOIN Maurice		FRANCES Yves
	BERLAND Yvon		FRANCOIS Georges
	BERNARD Dominique		FUENTES Pierre
	BERNARD Jean-Louis		GABRIEL Bernard
	BERNARD Jean-Paul		GALINIER Louis
	BERNARD Pierre-Marie		GALLAIS Hervé
	BERTRAND Edmond		GAMERRE Marc
	BISSET Jean-Pierre		GARCIN Michel
	BLANC Bernard		GARNIER Jean-Marc
	BLANC Jean-Louis		GAUTHIER André
	BOLLINI Gérard		GERARD Raymond
	BONGRAND Pierre		GEROLAMI-SANTANDREA André
	BONNEAU Henri		GIUDICELLI Sébastien
	BONNOIT Jean		GOUDARD Alain
	BORY Michel		GOUIN François
	BOTTA Alain		GRILLO Jean-Marie
	BOTTA-FRIDLUND Danielle		GRIMAUD Jean-Charles
	BOUBLI Léon		GRISOLI François
	BOURGADE Augustin		GROULIER Pierre
	BOUVENOT Gilles		GUYS Jean-Michel
	BOUYALA Jean-Marie		HADIDA/SAYAG Jacqueline
	BREMOND Georges		HASSOUN Jacques
	BRICOT René		HEIM Marc
	BRUNET Christian		HOUEL Jean
	BUREAU Henri		HUGUET Jean-François
	CAMBOULIVES Jean		JAQUET Philippe
	CANNONI Maurice		JAMMES Yves
	CARTOUZOU Guy		JOUVE Paulette
	CAU Pierre		JUHAN Claude
	CHABOT Jean-Michel		JUIN Pierre
	CHAMLIAN Albert		KAPHAN Gérard
	CHARPIN Denis		KASBARIAN Michel
	CHARREL Michel		KLEISBAUER Jean-Pierre
	CHAUVEL Patrick		LACHARD Jean
	CHOUX Maurice		LAFFARGUE Pierre
	CIANFARANI François		LAUGIER René
	CLAVERIE Jean-Michel		LE TREUT Yves
	CLEMENT Robert		LEGRE Régis
	COMBALBERT André		LEVY Samuel
	CONTE-DEVOLX Bernard		LOUCHET Edmond
	CORRIOL Jacques		LOUIS René
	COULANGE Christian		LUCIANI Jean-Marie
	CURVALE Georges		MAGALON Guy
	DALMAS Henri		MAGNAN Jacques
	DE MICO Philippe		MALLAN- MANCINI Josette

## PROFESSEURS HONORAIRES

DELPERO Jean-Robert  
DESSEIN Alain  
DELARQUE Alain

MALMEJAC Claude  
MARANINCHI Dominique  
MARTIN Claude

## PROFESSEURS HONORAIRES

MM MATTEI Jean François  
MERCIER Claude  
MICHOTEY Georges  
MIRANDA François  
MONFORT Gérard  
MONGES André  
MONGIN Maurice  
MUNDLER Olivier  
NAZARIAN Serge  
NICOLI René  
NOIRCLERC Michel  
OLMER Michel  
OREHEK Jean  
PAPY Jean-Jacques  
PAULIN Raymond  
PELOUX Yves  
PENAUD Antony  
PENE Pierre  
PIANA Lucien  
PICAUD Robert  
PIGNOL Fernand  
POGGI Louis  
POITOUT Dominique  
PONCET Michel  
POUGET Jean  
PRIVAT Yvan  
QUILICHINI Francis  
RANQUE Jacques  
RANQUE Philippe  
RAOULT Didier  
RICHAUD Christian  
RIDINGS Bernard  
ROCHAT Hervé  
ROHNER Jean-Jacques  
ROUX Hubert  
ROUX Michel  
RUFO Marcel  
SAHEL José  
SALAMON Georges  
SALDUCCI Jacques  
SAMBUC Roland  
SAN MARCO Jean-Louis  
SANKALE Marc  
SARACCO Jacques  
SARLES Jacques  
SARLES - PHILIP Nicole  
SASTRE Bernard  
SCHIANO Alain  
SCOTTO Jean-Claude  
SEBAHOUN Gérard  
SEITZ Jean-François  
SERMENT Gérard  
SOULAYROL René  
TAMALET Jacques  
TARANGER-CHARPIN Colette  
THIRION Xavier  
THOMASSIN Jean-Marc

UNAL Daniel  
VAGUE Philippe  
VAGUE/JUHAN Irène  
VANUXEM Paul  
VERVLOET Daniel  
VIALETTES Bernard  
WEILLER Pierre-Jean



## EMERITAT

### 2008

M. le Professeur	LEVY Samuel	31/08/2011
Mme le Professeur	JUHAN-VAGUE Irène	31/08/2011
M. le Professeur	PONCET Michel	31/08/2011
M. le Professeur	KASBARIAN Michel	31/08/2011
M. le Professeur	ROBERTOUX Pierre	31/08/2011

### 2009

M. le Professeur	DJIANE Pierre	31/08/2011
M. le Professeur	VERVLOET Daniel	31/08/2012

### 2010

M. le Professeur	MAGNAN Jacques	31/12/2014
------------------	----------------	------------

### 2011

M. le Professeur	DI MARINO Vincent	31/08/2015
M. le Professeur	MARTIN Pierre	31/08/2015
M. le Professeur	METRAS Dominique	31/08/2015

### 2012

M. le Professeur	AUBANIAC Jean-Manuel	31/08/2015
M. le Professeur	BOUVENOT Gilles	31/08/2015
M. le Professeur	CAMBOULIVES Jean	31/08/2015
M. le Professeur	FAVRE Roger	31/08/2015
M. le Professeur	MATTEI Jean-François	31/08/2015
M. le Professeur	OLIVER Charles	31/08/2015
M. le Professeur	VERVLOET Daniel	31/08/2015

### 2013

M. le Professeur	BRANCHEREAU Alain	31/08/2016
M. le Professeur	CARAYON Pierre	31/08/2016
M. le Professeur	COZZONE Patrick	31/08/2016
M. le Professeur	DELMONT Jean	31/08/2016
M. le Professeur	HENRY Jean-François	31/08/2016
M. le Professeur	LE GUICHAOUA Marie-Roberte	31/08/2016
M. le Professeur	RUFO Marcel	31/08/2016
M. le Professeur	SEBAHOUN Gérard	31/08/2016

### 2014

M. le Professeur	FUENTES Pierre	31/08/2017
M. le Professeur	GAMERRE Marc	31/08/2017
M. le Professeur	MAGALON Guy	31/08/2017
M. le Professeur	PERAGUT Jean-Claude	31/08/2017
M. le Professeur	WEILLER Pierre-Jean	31/08/2017

### 2015

M. le Professeur	COULANGE Christian	31/08/2018
M. le Professeur	COURAND François	31/08/2018
M. le Professeur	FAVRE Roger	31/08/2016
M. le Professeur	MATTEI Jean-François	31/08/2016
M. le Professeur	OLIVER Charles	31/08/2016
M. le Professeur	VERVLOET Daniel	31/08/2016

## EMERITAT

### 2016

M. le Professeur	BONGRAND Pierre	31/08/2019
M. le Professeur	BOUVENOT Gilles	31/08/2017
M. le Professeur	BRUNET Christian	31/08/2019
M. le Professeur	CAU Pierre	31/08/2019
M. le Professeur	COZZONE Patrick	31/08/2017
M. le Professeur	FAVRE Roger	31/08/2017
M. le Professeur	FONTES Michel	31/08/2019
M. le Professeur	JAMMES Yves	31/08/2019
M. le Professeur	NAZARIAN Serge	31/08/2019
M. le Professeur	OLIVER Charles	31/08/2017
M. le Professeur	POITOUT Dominique	31/08/2019
M. le Professeur	SEBAHOUN Gérard	31/08/2017
M. le Professeur	VIALETTES Bernard	31/08/2019

### 2017

M. le Professeur	ALESSANDRINI Pierre	31/08/2020
M. le Professeur	BOUVENOT Gilles	31/08/2018
M. le Professeur	CHAUVEL Patrick	31/08/2020
M. le Professeur	COZZONE Pierre	31/08/2018
M. le Professeur	DELMONT Jean	31/08/2018
M. le Professeur	FAVRE Roger	31/08/2018
M. le Professeur	OLIVER Charles	31/08/2018
M. le Professeur	SEBBAHOUN Gérard	31/08/2018

### 2018

M. le Professeur	MARANINCHI Dominique	31/08/2021
M. le Professeur	BOUVENOT Gilles	31/08/2019
M. le Professeur	COZZONE Pierre	31/08/2019
M. le Professeur	DELMONT Jean	31/08/2019
M. le Professeur	FAVRE Roger	31/08/2019
M. le Professeur	OLIVER Charles	31/08/2019
M. le Professeur	RIDINGS Bernard	31/08/2021

### 2019

M. le Professeur	BERLAND Yvon	31/08/2022
M. le Professeur	CHARPIN Denis	31/08/2022
M. le Professeur	CLAVERIE Jean-Michel	31/08/2022
M. le Professeur	FRANCES Yves	31/08/2022
M. le Professeur	CAU Pierre	31/08/2020
M. le Professeur	COZZONE Patrick	31/08/2020
M. le Professeur	DELMONT Jean	31/08/2020
M. le Professeur	FAVRE Roger	31/08/2020
M. le Professeur	FONTES Michel	31/08/2020
M. le Professeur	MAGALON Guy	31/08/2020
M. le Professeur	NAZARIAN Serge	31/08/2020
M. le Professeur	OLIVER Charles	31/08/2020
M. le Professeur	WEILLER Pierre-Jean	31/08/2020

### 2020

M. le Professeur	DELPERO Jean-Robert	31/08/2023
M. le Professeur	GRIMAUD Jean-Charles	31/08/2023
M. le Professeur	SAMBUC Roland	31/08/2023
M. le Professeur	SEITZ Jean-François	31/08/2023
M. le Professeur	BERLAND Yvon	31/08/2022
M. le Professeur	CHARPIN Denis	31/08/2022
M. le Professeur	CLAVERIE Jean-Michel	31/08/2022
M. le Professeur	FRANCES Yves	31/08/2022
M. le Professeur	BONGRAND Pierre	31/08/2021
M. le Professeur	COZZONE Patrick	31/08/2021

## EMERITAT

M. le Professeur	FAVRE Roger	31/08/2021
M. le Professeur	FONTES Michel	31/08/2021
M. le Professeur	NAZARIAN Serge	31/08/2021
M. le Professeur	WEILLER Pierre-Jean	31/08/2021

2021

M. le Professeur	BOUBLI Léon	31/08/2024
M. le Professeur	LEGRE Régis	31/08/2024
M. le Professeur	RAOULT Didier	31/08/2024
M. le Professeur	DELPERO Jean-Robert	31/08/2023
M. le Professeur	GRIMAUD Jean-Charles	31/08/2023
M. le Professeur	SAMBUC Roland	31/08/2023
M. le Professeur	SEITZ Jean-François	31/08/2023
M. le Professeur	BERLAND Yvon	31/08/2022
M. le Professeur	CHARPIN Denis	31/08/2022
M. le Professeur	CLAVERIE Jean-Michel	31/08/2022
M. le Professeur	FRANCES Yves	31/08/2022
M. le Professeur	BONGRAND Pierre	31/08/2022
M. le Professeur	BRUNET Christian	31/08/2022
M. le Professeur	COZZONE Patrick	31/08/2022
M. le Professeur	FAVRE Roger	31/08/2022
M. le Professeur	FONTES Michel	31/08/2022
M. le Professeur	NAZARIAN Serge	31/08/2022
M. le Professeur	OLIVER Charles	31/08/2022

<b>1967</b>	
MM. les Professeurs	DADI (Italie) CID DOS SANTOS (Portugal)
<b>1974</b>	
MM. les Professeurs	MAC ILWAIN (Grande-Bretagne) T.A. LAMBO (Suisse)
<b>1975</b>	
MM. les Professeurs	O. SWENSON (U.S.A.) Lord J.WALTON of DETCHANT (Grande-Bretagne)
<b>1976</b>	
MM. les Professeurs	P. FRANCHIMONT (Belgique) Z.J. BOWERS (U.S.A.)
<b>1977</b>	
MM. les Professeurs	C. GAJDUSEK-Prix Nobel (U.S.A.) C.GIBBS (U.S.A.) J. DACIE (Grande-Bretagne)
<b>1978</b>	
M. le Président	F. HOUPHOUET-BOIGNY (Côte d'Ivoire)
<b>1980</b>	
MM. les Professeurs	A. MARGULIS (U.S.A.) R.D. ADAMS (U.S.A.)
<b>1981</b>	
MM. les Professeurs	H. RAPPAPORT (U.S.A.) M. SCHOU (Danemark) M. AMENT (U.S.A.) Sir A. HUXLEY (Grande-Bretagne) S. REFSUM (Norvège)
<b>1982</b>	
M. le Professeur	W.H. HENDREN (U.S.A.)
<b>1985</b>	
MM. les Professeurs	S. MASSRY (U.S.A.) KLINSMANN (R.D.A.)
<b>1986</b>	
MM. les Professeurs	E. MIHICH (U.S.A.) T. MUNSAT (U.S.A.) LIANA BOLIS (Suisse) L.P. ROWLAND (U.S.A.)
<b>1987</b>	
M. le Professeur	P.J. DYCK (U.S.A.)
<b>1988</b>	
MM. les Professeurs	R. BERGUER (U.S.A.) W.K. ENGEL (U.S.A.) V. ASKANAS (U.S.A.) J. WEHSTER KIRKLIN (U.S.A.) A. DAVIGNON (Canada) A. BETTARELLO (Brésil)
<b>1989</b>	
M. le Professeur	P. MUSTACCHI (U.S.A.)

<b>1990</b>	
MM. les Professeurs	J.G. MC LEOD (Australie) J. PORTER (U.S.A.)
<b>1991</b>	
MM. les Professeurs	J. Edward MC DADE (U.S.A.) W. BURGDORFER (U.S.A.)
<b>1992</b>	
MM. les Professeurs	H.G. SCHWARZACHER (Autriche) D. CARSON (U.S.A.) T. YAMAMURO (Japon)
<b>1994</b>	
MM. les Professeurs	G. KARPATI (Canada) W.J. KOLFF (U.S.A.)
<b>1995</b>	
MM. les Professeurs	D. WALKER (U.S.A.) M. MULLER (Suisse) V. BONOMINI (Italie)
<b>1997</b>	
MM. les Professeurs	C. DINARELLO (U.S.A.) D. STULBERG (U.S.A.) A. MEIKLE DAVISON (Grande-Bretagne) P.I. BRANEMARK (Suède)
<b>1998</b>	
MM. les Professeurs	O. JARDETSKY (U.S.A.)
<b>1999</b>	
MM. les Professeurs	J. BOTELLA LLUSIA (Espagne) D. COLLEN (Belgique) S. DIMAURO (U. S. A.)
<b>2000</b>	
MM. les Professeurs	D. SPIEGEL (U. S. A.) C. R. CONTI (U.S.A.)
<b>2001</b>	
MM. les Professeurs	P-B. BENNET (U. S. A.) G. HUGUES (Grande Bretagne) J-J. O'CONNOR (Grande Bretagne)
<b>2002</b>	
MM. les Professeurs	M. ABEDI (Canada) K. DAI (Chine)
<b>2003</b>	
M. le Professeur Sir	T. MARRIE (Canada) G.K. RADDI (Grande Bretagne)
<b>2004</b>	
M. le Professeur	M. DAKE (U.S.A.)
<b>2005</b>	
M. le Professeur	L. CAVALLI-SFORZA (U.S.A.)
<b>2006</b>	
M. le Professeur	A. R. CASTANEDA (U.S.A.)

**2007**

M. le Professeur

S. KAUFMANN (Allemagne)

PROFESSEURS DES UNIVERSITES-PRATICIENS HOSPITALIERS

AGOSTINI FERRANDES Aubert	CHIARONI Jacques
<i>ALBANESE Jacques Surnombre</i>	CHINOT Olivier
ALIMI Yves	CHOSSEGROS Cyrille
AMABILE Philippe	COLLART Frédéric
AMBROSI Pierre	COSTELLO Régis
ANDRE Nicolas	COURBIERE Blandine
ARGENSON Jean-Noël	COWEN Didier
ASTOUL Philippe	CRAVELLO Ludovic
ATTARIAN Shahram	CUISSET Thomas
AUDOUIN Bertrand	DA FONSECA David
AUQUIER Pascal	DAHAN-ALCARAZ Laetitia
AVIERINOS Jean-François	DANIEL Laurent
AZULAY Jean-Philippe	DARMON Patrice
BAILLY Daniel	DAUMAS Aurélie
BARLESI Fabrice	DAVID Thierry
BARLIER-SETTI Anne	D'ERCOLE Claude
BARLOGIS Vincent	D'JOURNO Xavier
BARTHET Marc	DEHARO Jean-Claude
BARTOLI Christophe	DELAPORTE Emmanuel
BARTOLI Jean-Michel	DENIS Danièle
BARTOLI Michel	DISDIER Patrick
BARTOLOMEI Fabrice	DODDOLI Christophe
BASTIDE Cyrille	DRANCOURT Michel
BELIARD-LASSERRE Sophie	DUBUS Jean-Christophe
BENSOUSSAN Laurent	DUFFAUD Florence
BERBIS Philippe	DUFOUR Henry
BERBIS Julie	DURAND Jean-Marc
BERDAH Stéphane	DUSSOL Bertrand
BEROUD Christophe	EBBO Mikaël
BERTRAND Baptiste	EUSEBIO Alexandre
BERTUCCI François	FABRE Alexandre
BEYER-BERJOT Laura	FAKHRY Nicolas
BLAISE Didier	FAURE Alice
BLIN Olivier	FELICIAN Olivier
BLONDEL Benjamin	FENOLLAR Florence
BOISSIER Romain	FIGARELLA/BRANGER Dominique
BONIN/GUILLAUME Sylvie	FLECHER Xavier
BONELLO Laurent	FOUILLOUX Virginie
BONNET Jean-Louis	FOURNIER Pierre-Edouard
BOUFI Mourad	FRANCESCHI Frédéric
BOYER Laurent	FUENTES Stéphane
BREGÉON Fabienne	GABERT Jean
BRETELLE Florence	GABORIT Bénédicte
BROUQUI Philippe	GAINNIER Marc
BRUDER Nicolas	GARCIA Stéphane
BRUE Thierry	GARIBOLDI Vlad
BRUNET Philippe	GAUDART Jean
BURTEY Stéphane	GAUDY-MARQUESTE Caroline
CARCOPINO-TUSOLI Xavier	GENTILE Stéphanie
CASANOVA Dominique	GERBEAUX Patrick
CASTINETTI Frédéric	GEROLAMI/SANTANDREA René
CECCALDI Mathieu	GILBERT/ALESSI Marie-Christine
CHAGNAUD Christophe	GIORGI Roch
CHAMBOST Hervé	GIOVANNI Antoine
CHAMPSAUR Pierre	GIRARD Nadine
CHANEZ Pascal	GIRAUD/CHABROL Brigitte
CHARAFFE-JAUFFRET Emmanuelle	GONCALVES Anthony
CHARREL Rémi	GRANEL/REY Brigitte
CHAUMOITRE Kathia	GRANDVAL Philippe
	GREILLIER Laurent

PROFESSEURS DES UNIVERSITES-PRATICIENS HOSPITALIERS

AMABILE Philippe	COLLART Frédéric	GUIS Sandrine
AMBROSI Pierre	COSTELLO Régis	GUYE Maxime
ANDRE Nicolas	COURBIERE Blandine	GUYOT Laurent
ARGENSON Jean-Noël	COWEN Didier	<i>GUY'S Jean-Michel Retraite le 24/09/2021</i>
ASTOUL Philippe	CRAVELLO Ludovic	HABIB Gilbert
ATTARIAN Shahram	CUISSET Thomas	HARDWIGSEN Jean
AUDOUIN Bertrand	DA FONSECA David	HARLE Jean-Robert
AUQUIER Pascal	DAHAN-ALCARAZ Laetitia	HOUVENAEGHEL Gilles
AVIERINOS Jean-François	DANIEL Laurent	JACQUIER Alexis
AZULAY Jean-Philippe	DARMON Patrice	JOURDE-CHICHE Noémie
BAILLY Daniel	DAUMAS Aurélie	JOUBE Jean-Luc
BARLESI Fabrice	DAVID Thierry	KAPLANSKI Gilles
BARLIER-SETTI Anne	D'ERCOLE Claude	KARSENTY Gilles
BARLOGIS Vincent	D'JOURNO Xavier	<i>KERBAUL François détachement</i>
BARTHET Marc	DEHARO Jean-Claude	KRAHN Martin
BARTOLI Christophe	DELAPORTE Emmanuel	LAFFORGUE Pierre
BARTOLI Jean-Michel	DENIS Danièle	LAGIER Jean-Christophe
BARTOLI Michel	DISDIER Patrick	LAMBAUDIE Eric
BARTOLOMEI Fabrice	DODDOLI Christophe	LANCON Christophe
BASTIDE Cyrille	DRANCOURT Michel	LA SCOLA Bernard
BELIARD-LASSERRE Sophie	DUBUS Jean-Christophe	LAUNAY Franck
BENSOUSSAN Laurent	DUFFAUD Florence	LAVIEILLE Jean-Pierre
BERBIS Philippe	DUFOUR Henry	LE CORROLLER Thomas
BERBIS Julie	DURAND Jean-Marc	LECHEVALLIER Eric
BERDAH Stéphane	DUSSOL Bertrand	LEHUCHER-MICHEL Marie-Pascale
BEROUD Christophe	EBBO Mikael	LEONE Marc
BERTRAND Baptiste	EUSEBIO Alexandre	LEONETTI Georges
BERTUCCI François	FABRE Alexandre	LEPIDI Hubert
BEYER-BERJOT Laura	FAKHRY Nicolas	LEVY Nicolas
BLAISE Didier	FAURE Alice	MACE Loïc
BLIN Olivier	FELICIAN Olivier	MAGNAN Pierre-Edouard
BLONDEL Benjamin	FENOLLAR Florence	MANCINI Julien
BOISSIER Romain	FIGARELLA/BRANGER Dominique	MEGE Jean-Louis
BONIN/GUILLAUME Sylvie	FLECHER Xavier	MERROT Thierry
BONELLO Laurent	FOUILLOUX Virginie	METZLER/GUILLEMAIN Catherine
BONNET Jean-Louis	FOURNIER Pierre-Edouard	MEYER/DUTOUR Anne
BOUFI Mourad	FRANCESCHI Frédéric	MICCALEF/ROLL Joëlle
BOYER Laurent	FUENTES Stéphane	MICHEL Fabrice
BREGEON Fabienne	GABERT Jean	MICHEL Gérard
BRETELLE Florence	GABORIT Bénédicte	MICHEL Justin
BROUQUI Philippe	GAINNIER Marc	MICHELET Pierre
BRUDER Nicolas	GARCIA Stéphane	MILH Mathieu
BRUE Thierry	GARIBOLDI Vlad	MILLION Matthieu
BRUNET Philippe	GAUDART Jean	MOAL Valérie
BURTEY Stéphane	GAUDY-MARQUESTE Caroline	MORANGE Pierre-Emmanuel
CARCOPINO-TUSOLI Xavier	GENTILE Stéphanie	MOULIN Guy
CASANOVA Dominique	GERBEAUX Patrick	MOUTARDIER Vincent
CASTINETTI Frédéric	GEROLAMI/SANTANDREA René	NAUDIN Jean
CECCALDI Mathieu	GILBERT/ALESSI Marie-Christine	NICOLAS DE LAMBALLERIE Xavier
CHAGNAUD Christophe	GIORGI Roch	NICOLLAS Richard
CHAMBOST Hervé	GIOVANNI Antoine	NGUYEN Karine
CHAMPSAUR Pierre	GIRARD Nadine	OLIVE Daniel
CHANEZ Pascal	GIRAUD/CHABROL Brigitte	OLLIVIER Matthieu
CHARAFFE-JAUFFRET Emmanuelle	GONCALVES Anthony	OUAFIK L'Houcine
CHARREL Rémi	GRANEL/REY Brigitte	OVAERT-REGGIO Caroline
CHAUMOITRE Kathia	GRANDVAL Philippe	PAGANELLI Franck
	GREILLIER Laurent	<i>PANUEL Michel Surnombre</i>



PROFESSEURS DES UNIVERSITES-PRATICIENS HOSPITALIERS

PAPAZIAN Laurent	ROLL Patrice	VALERO René
PAROLA Philippe	ROSSI Dominique	VAROQUAUX Arthur Damien
PELISSIER-ALICOT Anne-Laure	ROSSI Pascal	VELLY Lionel
PELLETIER Jean	ROUDIER Jean	VEY Norbert
PERRIN Jeanne	SALAS Sébastien	VIDAL Vincent
PESENTI Sébastien	SARLON-BARTOLI Gabrielle	VIENS Patrice
PETIT Philippe	SCAVARDA Didier	VILLANI Patrick
PHAM Thao	SCHLEINITZ Nicolas	VITON Jean-Michel
PIERCECCHI/MARTI Marie-Dominique	SEBAG Frédéric	VITTON Véronique
PIQUET Philippe	SIELEZNEFF Igor	<i>VIEHWEGER Heide Elke détachement</i>
PIRRO Nicolas	SIMON Nicolas	VIVIER Eric
POINSO François	STEIN Andréas	XERRI Luc
RACCAH Denis	SUISSA Laurent	ZIELEZKIEWICZ Laurent
RANQUE Stéphane	TAIEB David	
REGIS Jean	THOMAS Pascal	
REYNAUD/GAUBERT Martine	THUNY Franck	
REYNAUD Rachel	TOSELLO Barthélémy	
RICHARD/LALLEMAND Marie-Alet	TREBUCHON-DA FONSECA Agnès	
RICHERI Raphaëlle	<i>TRIGLIA Jean-Michel Surnombre</i>	
ROCHE Pierre-Hugues	TROPIANO Patrick	
ROCH Antoine	TSIMARATOS Michel	
ROCHWERGER Richard	TURRINI Olivier	

**PROFESSEUR DES UNIVERSITES**

ADALIAN Pascal  
AGHABABIAN Valérie  
BELIN Pascal  
CHABANNON Christian  
CHABRIERE Eric  
FERON François  
LE COZ Pierre  
LEVASSEUR Anthony  
RANJEVA Jean-Philippe  
SOBOL Hagay

**PROFESSEUR CERTIFIE**

FRAISSE-MANGIALOMINI Jeanne

**PROFESSEUR DES UNIVERSITES ASSOCIE à MI-TEMPS**

REVIS Joana

**PROFESSEUR DES UNIVERSITES MEDECINE GENERALE**

GENTILE Gaëtan

**PROFESSEUR DES UNIVERSITES ASSOCIE à TEMPS PLEIN DES DISCIPLINES MEDICALES**

BOUSSUGES Alain

**PROFESSEUR DES UNIVERSITES ASSOCIE à MI-TEMPS DES DISCIPLINES MEDICALES**

DUTAU Hervé

# MAITRES DE CONFERENCES DES UNIVERSITES-PRATICIENS HOSPITALIERS

AHERFI Sarah	GELSI/BOYER Véronique	ROBERT Thomas
ANGELAKIS Emmanouil ( <i>disponibilité</i> )	GIUSIANO Bernard	ROMANET Pauline
ATLAN Catherine ( <i>disponibilité</i> )	GIUSIANO COURCAMBECK Sophie	SABATIER Renaud
BEGE Thierry	GONZALEZ Jean-Michel	SARI-MINODIER Irène
BENYAMINE Audrey	GOURIET Frédérique	SAVEANU Alexandru
BIRNBAUM David	GRAILLON Thomas	SECQ Véronique
BONINI Francesca	GUERIN Carole	SOLER Raphael
BOUCRAUT Joseph	GUENOUN MEYSSIGNAC Daphné	STELLMANN Jan-Patrick
BOULAMERY Audrey	GUIDON Catherine	SUCHON Pierre
BOULLU/CIOCCA Sandrine	GUIVARCH Jokthan	TABOURET Emeline
BOUSSEN Salah Michel	HAUTIER Aurélie	TOGA Isabelle
BUFFAT Christophe	HRAIECH Sami	TOMASINI Pascale
CAMILLERI Serge	IBRAHIM KOSTA Manal	TROUDE Lucas
CARRON Romain	JALOUX Charlotte	TROUSSE Delphine
CASSAGNE Carole	JARROT Pierre-André	TUCHTAN-TORRENTS Lucile
CERMOLACCE Michel	KASPI-PEZZOLI Elise	VELY Frédéric
CHAUDET Hervé	L'OLLIVIER Coralie	VENTON Geoffroy
CHRETIEN Anne-Sophie	LABIT-BOUVIER Corinne	VION-DURY Jean
COZE Carole	LAFAGE/POCHITALOFF-HUVALE Marina	ZATTARA/CANNONI Hélène
CUNY Thomas	LAGARDE Stanislas	
DADOUN Frédéric ( <i>disponibilité</i> )	LAGIER Aude ( <i>disponibilité</i> )	
DALES Jean-Philippe	LAGOUANELLE/SIMEONI Marie-Claude	
DARIEL Anne	LAMBERT Isabelle	
DEGEORGES/VITTE Joëlle ( <i>disponibilité</i> )	LENOIR Marien	
DEHARO Pierre	LEVY/MOZZICONACCI Annie	
DELLIAUX Stéphane	LOOSVELD Marie	
DELTEIL Clémence	MAAROUF Adil	
DESPLAT/JEGO Sophie	MACAGNO Nicolas	
DEVILLIER Raynier	MARLINGE Marion	
DUBOURG Grégory	MAUES DE PAULA André	
DUCONSEIL Pauline	MOTTOLA GHIGO Giovanna	
DUFOUR Jean-Charles	MEGE Diane	
ELDIN Carole	MOTTOLA GHIGO Giovanna	
FOLETTI Jean- Marc	NOUGAIREDE Antoine	
FRANKEL Diane	PAULMYER/LACROIX Odile	
FROMNOT Julien	RADULESCO Thomas	
GASTALDI Marguerite	RESSEGUIER Noémie	
GAUDRY Marine	ROBERT Philippe	

## MAITRES DE CONFERENCES DES UNIVERSITES

(mono-appartenants)

ABU ZAINEH Mohammad	DESNUES Benoît	RUEL Jérôme
BARBACARU/PERLES T. A.	MARANINCHI Marie	THOLLON Lionel
BERLAND Caroline	MERHEJ/CHAUVEAU Vicky	THIRION Sylvie
BOYER Sylvie	MINVIELLE/DEVICTOR Bénédicte	VERNA Emeline
COLSON Sébastien	POGGI Marjorie	
DEGIOANNI/SALLE Anna	POUGET Benoît	

## MAITRE DE CONFERENCES DES UNIVERSITES DE MEDECINE GENERALE

CASANOVA Ludovic

**MAITRES DE CONFERENCES ASSOCIES DE  
MEDECINE GENERALE à MI-TEMPS**

BARGIER Jacques  
FIERLING Thomas  
FORTE Jenny  
JANCZEWSKI Aurélie  
NUSSLI Nicolas  
ROUSSEAU-DURAND Raphaëlle  
THERY Didier

**MAITRE DE CONFERENCES ASSOCIE à MI-TEMPS**

BOURRIQUEN Maryline  
EVANS-VIALLAT Catherine  
LAZZAROTTO Sébastien  
LUCAS Guillaume  
MATHIEU Marion  
MAYENS-RODRIGUES Sandrine  
MELLINAS Marie  
ROMAN Christophe  
TRINQUET Laure

**PROFESSEURS DES UNIVERSITES et MAITRES DE CONFERENCES DES UNIVERSITES -  
PRATICIENS HOSPITALI PROFESSEURS ASSOCIES, MAITRES DE CONFERENCES DES  
UNIVERSITES mono-appartenants**

**ANATOMIE 4201**

CHAMPSAUR Pierre (PU-PH)  
LE CORROLLER Thomas (PU-PH)  
PIRRO Nicolas (PU-PH)

GUENOUN-MEYSSIGNAC Daphné (MCU-PH)  
LAGIER Aude (MCU-PH) *disponibilité*

THOLLON Lionel (MCF) (60ème section)

**ANATOMIE ET CYTOLOGIE PATHOLOGIQUES 4203**

CHARAFE/JAUFFRET Emmanuelle (PU-PH)  
DANIEL Laurent (PU-PH)  
FIGARELLA/BRANGER Dominique (PU-PH)  
GARCIA Stéphane (PU-PH)  
XERRI Luc (PU-PH)

DALES Jean-Philippe (MCU-PH)  
GIUSIANO COURCAMBECK Sophie (MCU PH)  
LABIT/BOUVIER Corinne (MCU-PH)  
MACAGNO Nicolas (MCU-PH)  
MAUES DE PAULA André (MCU-PH)  
SECQ Véronique (MCU-PH)

**ANESTHESIOLOGIE ET REANIMATION CHIRURGICALE ;  
MEDECINE URGENCE 4801**

ALBANESE Jacques (PU-PH) *Surnombre*  
BRUDER Nicolas (PU-PH)  
LEONE Marc (PU-PH)  
MICHEL Fabrice (PU-PH)  
VELLY Lionel (PU-PH)  
ZIELEZKIEWICZ Laurent (PU-PH)

BOUSSEN Salah Michel (MCU-PH)  
GUIDON Catherine (MCU-PH)

**ANGLAIS 11**

FRAISSE-MANGIALOMINI Jeanne (PRCE)

**BIOLOGIE ET MEDECINE DU DEVELOPPEMENT  
ET DE LA REPRODUCTION ; GYNECOLOGIE MEDICALE 5405**

METZLER/GUILLEMAIN Catherine (PU-PH)  
PERRIN Jeanne (PU-PH)

**ANTHROPOLOGIE 20**

ADALIAN Pascal (PR)

DEGIOANNI/SALLE Anna (MCF)  
POUGET Benoît (MCF)  
VERNA Emeline (MCF)

**BACTERIOLOGIE-VIROLOGIE ; HYGIENE HOSPITALIERE 4501**

CHARREL Rémi (PU PH)  
DRANCOURT Michel (PU-PH)  
FENOLLAR Florence (PU-PH)  
FOURNIER Pierre-Edouard (PU-PH)  
NICOLAS DE LAMBALLERIE Xavier (PU-PH)  
LA SCOLA Bernard (PU-PH)

AHERFI Sarah (MCU-PH)  
ANGELAKIS Emmanouil (MCU-PH) *disponibilité*  
DUBOURG Grégory (MCU-PH)  
GOURIET Frédérique (MCU-PH)  
NOUGAIREDE Antoine (MCU-PH)  
NINOVE Laetitia (MCU-PH)

CHABRIERE Eric (PR) (64ème section)

LEVASSEUR Anthony (PR) (64ème section)  
DESNUES Benoit (MCF) ( 65ème section )  
MERHEJ/CHAUVEAU Vicky (MCF) (87ème section)

**BIOCHIMIE ET BIOLOGIE MOLECULAIRE 4401**

BARLIER/SETTI Anne (PU-PH)  
GABERT Jean (PU-PH)  
GUIEU Régis (PU-PH)  
OUAFIK L'Houcine (PU-PH)

BUFFAT Christophe (MCU-PH)  
FROMNOT Julien (MCU-PH)  
MARLINGE Marion (MCU-PH)  
MOTTOLA GHIGO Giovanna (MCU-PH)  
ROMANET Pauline (MCU-PH)  
SAVEANU Alexandru (MCU-PH)

**BIOLOGIE CELLULAIRE 4403**

ROLL Patrice (PU-PH)

FRANKEL Diane (MCU-PH)  
GASTALDI Marguerite (MCU-PH)  
KASPI-PEZZOLI Elise (MCU-PH)  
LEVY-MOZZICONNACCI Annie (MCU-PH)

**BIOPHYSIQUE ET MEDECINE NUCLEAIRE 4301**

GUEDJ Eric (PU-PH)  
 GUYE Maxime (PU-PH)  
 TAIEB David (PU-PH)

BELIN Pascal (PR) (69ème section)  
 RANJEVA Jean-Philippe (PR) (69ème section)

CAMMILLERI Serge (MCU-PH)  
 VION-DURY Jean (MCU-PH)

BARBACARU/PERLES Téodora Adriana (MCF) (69ème section)

**BIostatistiques, Informatique Médicale  
ET TECHNOLOGIES DE COMMUNICATION 4604**

GAUDART Jean (PU-PH)  
 GIORGI Roch (PU-PH)  
 MANCINI Julien (PU-PH)

CHAUDET Hervé (MCU-PH)  
 DUFOUR Jean-Charles (MCU-PH)  
 GIUSIANO Bernard (MCU-PH)

ABU ZAINEH Mohammad (MCF) (5ème section)  
 BOYER Sylvie (MCF) (5ème section)

**CHIRURGIE ORTHOPEDIQUE ET TRAUMATOLOGIQUE 5002**

ARGENSON Jean-Noël (PU-PH)  
 BLONDEL Benjamin (PU-PH)  
 FLECHER Xavier (PU PH)  
 OLLIVIER Matthieu (PU-PH)  
 ROCHWERGER Richard (PU-PH)  
 TROPIANO Patrick (PU-PH)

**CANCEROLOGIE ; RADIOTHERAPIE 4702**

BERTUCCI François (PU-PH)  
 CHINOT Olivier (PU-PH)  
 COWEN Didier (PU-PH)  
 DUFFAUD Florence (PU-PH)  
 GONCALVES Anthony (PU-PH)  
 HOUVENAEGHEL Gilles (PU-PH)  
 LAMBAUDIE Eric (PU-PH)  
 PADOVANI Laetitia (PH-PH)  
 SALAS Sébastien (PU-PH)  
 VIENS Patrice (PU-PH)

SABATIER Renaud (MCU-PH)  
 TABOURET Emeline (MCU-PH)

**CARDIOLOGIE 5102**

AVIERINOS Jean-François (PU-PH)  
 BONELLO Laurent (PU PH) BONNET Jean-Louis (PU-PH)  
 CUISSET Thomas (PU-PH) DEHARO Jean-Claude (PU-PH)  
 FRANCESCHI Frédéric (PU-PH) HABIB Gilbert (PU-PH)  
 PAGANELLI Franck (PU-PH) THUNY Franck (PU-PH)

DEHARO Pierre (MCU PH)

**CHIRURGIE VISCERALE ET DIGESTIVE 5202**

BERDAH Stéphane (PU-PH)  
 BEYER-BERJOT Laura (PU-PH)  
 HARDWIGSEN Jean (PU-PH)  
 MOUTARDIER Vincent (PU-PH)  
 SEBAG Frédéric (PU-PH)  
 SIELEZNEFF Igor (PU-PH)  
 TURRINI Olivier (PU-PH)

BEGE Thierry (MCU-PH)  
 BIRNBAUM David (MCU-PH)  
 DUCONSEIL Pauline (MCU-PH)  
 GUERIN Carole (MCU PH)  
 MEGE Diane (MCU-PH)

**CHIRURGIE INFANTILE 5402**

*GUYS Jean-Michel (PU-PH) Retraite le 24/09/2021*  
 FAURE Alice (PU PH)  
 JOUVE Jean-Luc (PU-PH)  
 LAUNAY Franck (PU-PH)  
 MERROT Thierry (PU-PH)  
 PESENTI Sébastien (PU-PH)  
 VIEHWEGER Heide Elke (PU-PH) détachement

DARIEL Anne (MCU-PH)

**CHIRURGIE MAXILLO-FACIALE ET STOMATOLOGIE 5503**

CHOSSEGROS Cyrille (PU-PH)  
 GUYOT Laurent (PU-PH)

FOLETTI Jean-Marc (MCU-PH)

**CHIRURGIE THORACIQUE ET CARDIOVASCULAIRE 5103**

COLLART Frédéric (PU-PH)  
 D'JOURNO Xavier (PU-PH)  
 DODDOLI Christophe (PU-PH)  
 FOUILLOUX Virginie (PU-PH)  
 GARIBOLDI Vlad (PU-PH)  
 MACE Loïc (PU-PH)  
 THOMAS Pascal (PU-PH)

LENOIR Marien (MCU-PH)  
 TROUSSE Delphine (MCU-PH)

**CHIRURGIE PLASTIQUE,  
RECONSTRUCTRICE ET ESTHETIQUE ; BRÛLOGIE 5004**

BERTRAND Baptiste (PU-PH)  
 CASANOVA Dominique (PU-PH)

HAUTIER Aurélie (MCU-PH)  
 JALOUX Charlotte (MCU PH)

**CHIRURGIE VASCULAIRE ; MEDECINE VASCULAIRE 5104**

ALIMI Yves (PU-PH)  
 AMABILE Philippe (PU-PH)  
 BARTOLI Michel (PU-PH)  
 BOUFI Mourad (PU-PH)  
 MAGNAN Pierre-Edouard (PU-PH)  
 PIQUET Philippe (PU-PH)  
 SARLON-BARTOLI Gabrielle (PU PH)

GAUDRY Marine (MCU PH)  
 SOLER Raphael (MCU-PH)

**GASTROENTEROLOGIE ; HEPATOLOGIE ; ADDICTOLOGIE 5201**

BARTHET Marc (PU-PH)  
 DAHAN-ALCARAZ Laetitia (PU-PH)  
 GEROLAMI-SANTANDREA René (PU-PH)  
 GRANDVAL Philippe (PU-PH)  
 VITTON Véronique (PU-PH)

**HISTOLOGIE, EMBRYOLOGIE ET CYTOGENETIQUE 4202**

LEPIDI Hubert (PU-PH)

PAULMYER/LACROIX Odile (MCU-PH)

GONZALEZ Jean-Michel (MCU-PH)

**GENETIQUE 4704****DERMATOLOGIE - VENEREOLOGIE 5003**

BERBIS Philippe (PU-PH)  
 DELAPORTE Emmanuel (PU-PH)  
 GAUDY/MARQUESTE Caroline (PU-PH)  
 GROB Jean-Jacques (PU-PH)  
 RICHARD/LALLEMAND Marie-Aleth (PU-PH)

BEROUD Christophe (PU-PH)  
 KRAHN Martin (PU-PH)  
 LEVY Nicolas (PU-PH)  
 NGYUEN Karine (PU-PH)  
  
 ZATTARA/CANNONI Hélène (MCU-PH)

**DUSI**

COLSON Sébastien (MCF)

BOURRIQUEN Maryline (MAST)  
 EVANS-VIALLAT Catherine (MAST)  
 LUCAS Guillaume (MAST)  
 MAYEN-RODRIGUES Sandrine (MAST)  
 MELLINAS Marie (MAST)  
 ROMAN Christophe (MAST)  
 TRINQUET Laure (MAST)

**GYNECOLOGIE-OBSTETRIQUE ; GYNECOLOGIE MEDICALE 5403**

AGOSTINI Aubert (PU-PH)  
 BRETTELLE Florence (PU-PH)  
 CARCOPINO-TUSOLI Xavier (PU-PH)  
 COURBIERE Blandine (PU-PH)  
 CRAVELLO Ludovic (PU-PH)  
 D'ERCOLE Claude (PU-PH)

**ENDOCRINOLOGIE ,DIABETE ET MALADIES METABOLIQUES ;  
GYNECOLOGIE MEDICALE 5404**

BRUE Thierry (PU-PH)  
 CASTINETTI Frédéric (PU-PH)

CUNY Thomas (MCU PH)

**EPIDEMIOLOGIE, ECONOMIE DE LA SANTE ET PREVENTION**

4601

AUQUIER Pascal (PU-PH)  
BERBIS Julie (PU-PH)  
BOYER Laurent (PU-PH)  
GENTILE Stéphanie (PU-PH)

LAGOUANELLE/SIMEONI Marie-Claude (MCU-PH)  
RESSEGUIER Noémie (MCU-PH)

MINVIELLE/DEVICTOR Bénédicte (MCF)(06ème section)

**IMMUNOLOGIE 4703I**

KAPLANSKI Gilles (PU-PH)  
MEGE Jean-Louis (PU-PH)  
OLIVE Daniel (PU-PH)  
VIVIER Eric (PU-PH)

FERON François (PR) (69ème section)

BOUCRAUT Joseph (MCU-PH)  
CHRETIEN Anne-Sophie (MCU PH)  
DEGEORGES/VITTE Joëlle (MCU-PH)  
DESPLAT/JEGO Sophie (MCU-PH)  
JARROT Pierre-André (MCU PH)  
ROBERT Philippe (MCU-PH)  
VELY Frédéric (MCU-PH)

**MALADIES INFECTIEUSES ; MALADIES TROPICALES 4503**

BROUQUI Philippe (PU-PH)  
LAGIER Jean-Christophe (PU-PH)  
MILLION Matthieu (PU-PH)  
PAROLA Philippe (PU-PH)  
STEIN Andréas (PU-PH)

ELDIN Carole (MCU-PH)

**MEDECINE D'URGENCE 4805**

GERBEAUX Patrick (PU PH)  
*KERBAUL François (PU-PH) détachement*  
MICHELET Pierre (PU-PH)

**MEDECINE INTERNE ; GERIATRIE ET BIOLOGIE DU  
VIEILLISSEMENT ; ADDICTOLOGIE 5301**

BONIN/GUILLAUME Sylvie (PU-PH)      BENYAMINE Audrey (MCU-PH)  
DISDIER Patrick (PU-PH)  
DURAND Jean-Marc (PU-PH)  
EBBO Mikael (PU-PH)  
GRANEL/REY Brigitte (PU-PH)  
HARLE Jean-Robert (PU-PH)  
ROSSI Pascal (PU-PH)  
SCHLEINITZ Nicolas (PU-PH)

**HEMATOLOGIE ; TRANSFUSION 4701**

BLAISE Didier (PU-PH)  
COSTELLO Régis (PU-PH)  
CHIARONI Jacques (PU-PH)  
GILBERT/ALESSI Marie-Christine (PU-PH)  
MORANGE Pierre-Emmanuel (PU-PH)  
VEY Norbert (PU-PH)

DEVILLIER Raynier (MCU PH)  
GELSI/BOYER Véronique (MCU-PH)  
IBRAHIM KOSTA Manal (MCU PH)  
LAFAGE/POCHITALOFF-HUVALE Marina (MCU-PH)  
LOOSVELD Marie (MCU-PH)  
SUCHON Pierre (MCU-PH)  
VENTON (MCU-PH)

POGGI Marjorie (MCF) (64ème section)

**MEDECINE LEGALE ET DROIT DE LA SANTE 4603**

BARTOLI Christophe (PU-PH)  
LEONETTI Georges (PU-PH)  
PELISSIER-ALICOT Anne-Laure (PU-PH)  
PIERCECCHI-MARTI Marie-Dominique (PU-PH)

DELTEIL Clémence (MCU PH)  
TUCHTAN-TORRENTS Lucile (MCU-PH)

BERLAND Caroline (MCF) (1ère section)

**MEDECINE PHYSIQUE ET DE READAPTATION 4905**

BENSOUSSAN Laurent (PU-PH)  
VITON Jean-Michel (PU-PH)

**MEDECINE ET SANTE AU TRAVAIL 4602**

LEHUCHER/MICHEL Marie-Pascale (PU-PH)

SARI/MINODIER Irène (MCU-PH)



**MEDECINE GENERALE 5303**

GENTILE Gaëtan (PR Méd. Gén. Temps plein)

CASANOVA Ludovic (MCF Méd. Gén. Temps plein)

BARGIER Jacques (MCF associé Méd. Gén. À mi-temps)

FIERLING Thomas (MCF associé Méd. Gén. À mi-temps)

FORTE Jenny (MCF associé Méd. Gén. À mi-temps)

JANCZEWSKI Aurélie (MCF associé Méd. Gén. À mi-temps)

NUSSLI Nicolas (MCF associé Méd. Gén. À mi-temps)

ROUSSEAU-DURAND Raphaëlle

(MCF associé Méd. Gén. À mi-temps)

THERY Didier (MCF associé Méd. Gén. À mi-temps)

(nomination au 1/10/2019)

**NUTRITION 4404**

BELIARD Sophie (PU-PH)

DARMON Patrice (PU-PH)

RACCAH Denis (PU-PH)

VALERO René (PU-PH)

*ATLAN Catherine (MCU-PH) disponibilité*

MARANINCHI Marie (MCF) (66ème section)

**ONCOLOGIE 65 (BIOLOGIE CELLULAIRE)**

CHABANNON Christian (PR) (66ème section)

SOBOL Hagay (PR) (65ème section)

**OPHTALMOLOGIE 5502**

DAVID Thierry (PU-PH)

DENIS Danièle (PU-PH)

**OTO-RHINO-LARYNGOLOGIE 5501**

DESSI Patrick (PU-PH)

FAKHRY Nicolas (PU-PH)

GIOVANNI Antoine (PU-PH)

LAVIEILLE Jean-Pierre (PU-PH)

MICHEL Justin (PU-PH)

NICOLLAS Richard (PU-PH)

*TRIGLIA Jean-Michel (PU-PH) Surnombre*

RADULESCO Thomas (MCU-PH)

REVIS Joana (PAST) (Orthophonie) (7ème Section)

**NEPHROLOGIE 5203**

BRUNET Philippe (PU-PH)

BURTEY Stéphanne (PU-PH)

DUSSOL Bertrand (PU-PH)

JOURDE CHICHE Noémie (PU PH)

MOAL Valérie (PU-PH)

ROBERT Thomas (MCU-PH)

**NEUROCHIRURGIE 4902**

DUFOUR Henry (PU-PH)

FUENTES Stéphane (PU-PH)

REGIS Jean (PU-PH)

ROCHE Pierre-Hugues (PU-PH)

SCAVARDA Didier (PU-PH)

CARRON Romain (MCU PH)

GRAILLON Thomas (MCU PH)

TROUDE Lucas (MCU-PH)

**NEUROLOGIE 4901**

ATTARIAN Sharham (PU PH)

AUDOIN Bertrand (PU-PH)

AZULAY Jean-Philippe (PU-PH)

CECCALDI Mathieu (PU-PH)

EUSEBIO Alexandre (PU-PH)

FELICIAN Olivier (PU-PH)

PELLETIER Jean (PU-PH)

SUISSA Laurent (PU-PH)

MAAROUF Adil (MCU-PH)

**PEDOPSYCHIATRIE; ADDICTOLOGIE 4904**

DA FONSECA David (PU-PH)

POINSO François (PU-PH)

GUIVARCH Jokthan (MCU-PH)

**PHARMACOLOGIE FONDAMENTALE -  
PHARMACOLOGIE CLINIQUE; ADDICTOLOGIE 4803**

BLIN Olivier (PU-PH)

MICALLEF/ROLL Joëlle (PU-PH)

SIMON Nicolas (PU-PH)

BOULAMERY Audrey (MCU-PH)

RANQUE Stéphane (PU-PH)

LE COZ Pierre (PR) (17ème section)

CASSAGNE Carole (MCU-PH)

MATHIEU Marion (MAST)

L'OLLIVIER Coralie (MCU-PH)

TOGA Isabelle (MCU-PH)

### PHYSIOLOGIE 4402

### PEDIATRIE 5401

ANDRE Nicolas (PU-PH)

BARTOLOMEI Fabrice (PU-PH)

BARLOGIS Vincent (PU-PH)

BREGEON Fabienne (PU-PH)

CHAMBOST Hervé (PU-PH)

GABORIT Bénédicte (PU-PH)

DUBUS Jean-Christophe (PU-PH)

MEYER/DUTOUR Anne (PU-PH)

FABRE Alexandre (PU-PH)

TREBUCHON/DA FONSECA Agnès (PU-PH)

GIRAUD/CHABROL Brigitte (PU-PH)

BOUSSUGES Alain (PR associé à temps plein)

MICHEL Gérard (PU-PH)

BONINI Francesca (MCU-PH)

MILH Mathieu (PU-PH)

BOULLU/CIOCCA Sandrine (MCU-PH)

OVAERT-REGGIO Caroline (PU-PH)

*DADOUN Frédéric (MCU-PH) (disponibilité)*

REYNAUD Rachel (PU-PH)

DELLIAUX Stéphane (MCU-PH)

TOSELLO Barthélémy (PU-PH)

LAGARDE Stanislas (MCU-PH)

TSIMARATOS Michel (PU-PH)

LAMBERT Isabelle (MCU-PH)

COZE Carole (MCU-PH)

RUEL Jérôme (MCF) (69ème section)

THIRION Sylvie (MCF) (66ème section)

### PSYCHIATRIE D'ADULTES ; ADDICTOLOGIE 4903

### PNEUMOLOGIE; ADDICTOLOGIE 5101

BAILLY Daniel (PU-PH)

ASTOUL Philippe (PU-PH)

LANCON Christophe (PU-PH)

BARLESI Fabrice (PU-PH)

NAUDIN Jean (PU-PH)

CHANEZ Pascal (PU-PH)

RICHERI Raphaëlle (PU-PH)

GREILLIER Laurent (PU PH)

CERMOLACCE Michel (MCU-PH)

REYNAUD/GAUBERT Martine (PU-PH)

### PSYCHOLOGIE - PSYCHOLOGIE CLINIQUE, PSYCHOLOGIE SOCIALE

TOMASINI Pascale (MCU-PH)

AGHABABIAN Valérie (PR)

DUTAU Hervé ( Pr Associé des universités à 1/2 temps)

LAZZAROTTO Sébastien (MAST)

### RADIOLOGIE ET IMAGERIE MEDICALE 4302

### RHUMATOLOGIE 5001

BARTOLI Jean-Michel (PU-PH)

GUIS Sandrine (PU-PH)

CHAGNAUD Christophe (PU-PH)

LAFFORGUE Pierre (PU-PH)

CHAUMOITRE Kathia (PU-PH)

PHAM Thao (PU-PH)

GIRARD Nadine (PU-PH)

ROUDIER Jean (PU-PH)

JACQUIER Alexis (PU-PH)

MOULIN Guy (PU-PH)

*PANUEL Michel (PU-PH) surnombre*

PETIT Philippe (PU-PH)

VAROQUAUX Arthur Damien (PU-PH)

VIDAL Vincent (PU-PH)

### THERAPEUTIQUE; MEDECINE D'URGENCE; ADDICTOLOGIE 4804

STELLMANN Jan-Patrick (MCU-PH)

AMBROSI Pierre (PU-PH)

DAUMAS Aurélie (PU-PH)

VILLANI Patrick (PU-PH)

GAINNIER Marc (PU-PH)  
PAPAZIAN Laurent (PU-PH)  
ROCH Antoine (PU-PH)

HRAIECH Sami (MCU-PH)

BASTIDE Cyrille (PU-PH)  
BOISSIER Romain (PU-PH)  
KARSENTY Gilles (PU-PH)  
LECHEVALLIER Eric (PU-PH)  
ROSSI Dominique (PU-PH)

## **Remerciement aux collaborateurs**

Je tenais à remercier le professeur Xavier Thirion d'avoir été mon mentor et de m'avoir lancé dans mon internat de Santé Publique.

Je remercie Anita Burgun, Antoine Neuraz et Ivan Lerner pour leur aide et accompagnement au cours de ce travail.

Merci à mon jury, le professeur Boyer, le professeur Berbis, le professeur Champsaur ainsi que le docteur Antonini d'avoir accepté d'assister à cette soutenance et de me permettre d'achever ce parcours.

## Remerciement personnels

À mes deux grands-pères, André et Jean-Pierre, cette thèse est pour vous. Merci d'avoir été si fiers.

À Oma, ma grand-mère bleue, tu as passé tant de temps à m'aider à m'épanouir, à développer ma créativité, à m'ouvrir au monde. Ce travail est aussi pour toi.

À Fanfan, ma douce grand-mère, qui est toujours là pour m'écouter, merci d'être encore à mes côtés pour cette étape de vie.

À mes parents, Éveline et Christophe, vous m'avez entouré d'amour et de valeurs justes, vous m'avez montré comment déplacer des montagnes, merci d'avoir autant participé à ce que je suis devenue. Je ne saurais comment vous exprimez tout mon amour et ma gratitude en quelques lignes, mais je continuerai à le faire chaque jour que j'aurais la chance de passer auprès de vous.

À mes beaux-parents, Anne-Isabelle et Bruno, vous êtes une richesse. Anne-Isabelle, je ne saurais te remercier de ta douceur et ta gentillesse. Merci d'aimer et de prendre soin de mon papa comme tu le fais. Bruno, mon médecin préféré toutes catégories, merci d'avoir été mon premier professeur, mentor, merci d'avoir été mon premier collègue, merci d'aimer ma maman comme tu le fais. Enfin merci de m'avoir donné le plus formidable des petits frères.

À mes frères et ma sœur, Robin, Marie et Gabriel.

À ma marraine, ma deuxième maman, toujours présente, douce et réconfortante.

À Jordi, merci d'avoir toujours cru en moi, merci d'être cet ami fidèle et si cher à mon cœur, tu m'as appris à m'adoucir et à grandir dans l'adversité. Merci pour ces années et pour les futurs qui nous attendent.

À Jérôme, mon camarade de sport, de discussion, de voyage. Mon ami.

À Frédéric, je suis obligé de remercier le sort qui plus d'une fois aura voulu que nous devenions amis. Tu auras été le début d'une nouvelle vie. Je remercie aussi tes parents, Catherine et François d'avoir si gentiment pris soin de moi au début de celle-ci. Ta constante amitié sans vague et faille est un cadeau.

À Antoine L, le seul et unique de mes amis que j'aurais mis des années à décrypter, challenge perpétuel et ami perpétuel, je ne peux que me réjouir de t'avoir à mes côtés.

À Alice, merci pour ces premiers bains de vitamine D que nous avons pris ensemble, une amie comme on n'en fait plus.

À Sabine, Faustine, Coralie, Albane, Sébastien, Anne-cécile, Antonin, mes amis de lycée.

À mes amis du SAIHM, Justin et Ahmed, merci de m'avoir fait confiance et d'avoir fait de mes

projets les nôtres.

À Sylvain, merci pour cette aventure au CLISP, cette collaboration aura été un moment incroyable d'apprentissage et d'amitié.

À Martin, mon "partner", je m'étonne de voir à quel point notre entente est harmonieuse, c'est une délectation journalière, bien que je ne puisse voir le futur j'ai tout de même le sentiment que ça sera pour longtemps.

À Thibaut, ta générosité et ta gentillesse te perdront, mais en attendant j'en profite ;).

À Davide, mon acolyte, mon italien, mon ami.

À Isabelle, ma party budy, ta présence est toujours une joie et une force. J'attends avec impatience le moment où la team infernale pourra se remettre en action.

À Myriam, "mimi", à ses nombreuses années qui nous reste à encore partager.

À Louisa, ma camarade de première année, rétrospectivement cette année fut formidable, et cela beaucoup grâce à toi et ton amitié.

À Antoine R, Madame Baronne de Bibiche tenait à te remercier pour ton amitié sans faille. Elle espère sincèrement pouvoir être aussi là pour toi encore très longtemps.

À Kelly, je te remercie d'avoir été mon soutien pendant cette première année de médecine, d'avoir pris soin de moi et de m'avoir accompagnée.

À Laure qui m'a poussée et aidée durant ces années d'externat et sans qui je ne serais surement pas allée au bout, je te remercie aussi de m'avoir encouragée à aller à Marseille, ville dans laquelle j'ai pu m'épanouir.

To Dagmar, you helped me get the last push I needed to finish those studies. Thank you so much. I look forward to building a life with you.

*"Le meilleur moyen de prédire l'avenir est de le créer"* Abraham Lincoln

# Sommaire

## 1 Introduction

## 2 Travaux connexes

## 3 Matériel et méthodes

### 3.1 Données et préparation des données

3.1.1. *Recueil des données*

3.1.2. *Préparation des données : CIM-10 (CIM-10 vers UMLS CUI)*

3.1.3. *Préparation des données : concepts des notes cliniques*

3.1.4. *Préparation des données : test de laboratoire (LOINC vers UMLS CUI)*

3.1.5. *Préparation des données : données démographiques*

3.1.6. *Vecteur patient*

### 3.2. Implémentation

3.2.1. *Architecture du modèle*

3.2.1.1. *Auto-encodeurs (AE)*

3.2.1.2. *Auto-encodeurs de débruitage (DAE)*

3.2.1.3. *Auto-encodeurs de débruitage superposés (SDAE)*

3.2.2. *Représentation réutilisable du patient : entraînement et hyperparamètres*

3.2.3. *Sous-tâches (réhospitalisation) : entraînement et hyperparamètres*

3.2.4. *Modèle de références (baselines)*

## 4 Résultats

### 4.1 Description de la population

4.1.1 *Données complètes*

4.1.2 *Données pédiatriques*

### 4.2 Préparation des données et mappage

4.2.1 *Données complètes*

4.2.2 *Données pédiatriques*

### 4.3 Phase d'apprentissage de l'auto-encodeur de débruitage superposés

### 4.4 Performances d'auto-encodeur de débruitages superposés

4.4.1 *Base de données complète*

4.4.2 *Données pédiatriques*

## 5 Discussion

## Bibliographie

**Annexe 1 : Présentation des architectures principal des réseaux de neurones**

**Annexe 2 : Présentation des limites de l'apprentissage des réseaux de neurones**

**Annexe 3 : Présentation détaillés de l'état de l'art**

**Annexe 4 : Table des figures et tableaux**

# Modélisation du dossier patient par deep learning : application un entrepôt de données cliniques français pédiatrique et de maladies rares

Mathilde Besson<sup>1</sup>

<sup>1</sup>Interne de Santé Publique, Aix Marseille Université

## Résumé

**Objectif** : Construire une modélisation du dossier patient par deep learning réutilisable pour différentes sous-tâches à partir de l'entrepôt de données cliniques de l'hôpital Necker puis l'appliquer à une sous-tâche de test : la réhospitalisation des patients en moins de 30 jours.

**Matériaux et méthodes** : Nous avons premièrement conçu un alignement (mapping) des codes CIM-10, des codes LOINC de biologie avec les CUI de l'UMLS. Dans un second temps nous avons utilisé les concepts précédemment extraits des notes médicales puis nous les avons alignés avec les CUI de l'UMLS. La modélisation a été construite avec une structure de deep learning semi-supervisée : les stacked denoising auto-encoder (SDAE). Nous avons testé le SDAE sur différents échantillons de la base de données (1000, 10000, toutes les données ainsi qu'un sous-échantillonnage).

**Résultats** : Le SDAE a obtenu de meilleurs résultats que l'ensemble des échantillons de références, à l'exception du sous-échantillonnage où le DEA a obtenu de meilleurs résultats.

**Discussion** : Nous montrons dans cette étude que la modélisation du dossier patient d'un hôpital pédiatrique par deep learning permet après application à une sous tâche de classification de prédire les patients réhospitalisés en moins de 30 jours. Plusieurs améliorations semblent envisageables et d'autres sous-tâches doivent être testées.

## Mots clés

Medical Informatics; Electronic health records; Deep learning; Unsupervised learning; Patient representations

## 1 Introduction

La prédiction d'événements en médecine est un enjeu important, que ce soit pour le diagnostic, les mouvements des patients ou leurs pronostic [1,2], en particulier dans le cas des maladies rares qui sont généralement difficiles à diagnostiquer en raison de leurs caractéristiques inhabituelles impliquant de longs retards de diagnostic. Ces retards de diagnostic ont des conséquences dramatiques pour les patients en termes de pronostic et de traitement. [3,4]. La représentation et la modélisation des données des dossiers patient informatisé (DPI) constituent un défi majeur, principalement en raison de leurs caractères bruités, hétérogènes, épars, irréguliers et biaisés [5]. Les observations sont également épisodiques et irrégulières dans le temps. Les dossiers contiennent des données hétérogènes telles que des notes cliniques, des codes de diagnostic, des procédures, des résultats de tests de laboratoire, etc. Leur intégration dans un modèle est essentielle pour



améliorer la prise en charge de certaines maladies rares. L'hôpital Necker accueille l'Institut Imagine, un institut de recherche sur les maladies génétiques. Ses bases de données particulières sont l'occasion d'explorer l'efficacité de modèles de deep learning déjà connus. Les difficultés de diagnostic pourraient être améliorées grâce aux aides innovantes récemment développées avec des modèles de deep learning utilisant les données des dossiers patients informatisés (DPI). Les modèles de deep learning sont développés afin de remplir différentes tâches, parmi lesquelles la prédiction des résultats, le phénotypage computationnel, l'extraction de connaissances et l'apprentissage de représentations. Construire une représentation réutilisable du patient, comme Miotto & al [5] en deep learning est la première étape afin de pouvoir l'appliquer à différentes sous-tâches.

Ainsi, à l'instar du plongement lexical (*word embedding*), l'apprentissage de la représentation des patients vise à créer un plongement des patients (*patient embedding*), c'est-à-dire une technique permettant de représenter les patients sous forme de vecteurs, cette représentation permettant d'obtenir des caractéristiques plus condensées et plus informatives qui peuvent par la suite être réutilisées pour remplir des tâches secondaires. Les algorithmes de deep learning fonctionnent bien sur de grandes bases de données, alors que les maladies rares sont, par nature, généralement représentées par de petites bases de données. L'utilisation d'une méthode auto-supervisée pourrait permettre une meilleure exploitation de ces données. Miotto & al. [5], dans le modèle Deep Patient, apprennent une représentation du patient et l'utilisent pour remplir une tâche de prédiction. Notre objectif principal est d'appliquer le modèle Deep Patient à l'entrepôt de données de l'hôpital Necker et de tester une sous-tâche : la prédiction des réhospitalisations des patients en moins de 30 jours.

## 2 Travaux connexes

Il y a moins de 30 ans, l'apprentissage profond a été introduit dans la communauté du *machine learning* [6], depuis il a été appliqué à de nombreux types de données avec de nombreuses architectures. Dans le domaine des soins de santé, la dernière décennie a vu une explosion de la quantité de données stockées dans les dossiers patients informatisés (DPI), par conséquent les applications des techniques de *deep learning* se sont développées [7].

Les techniques de *deep learning* sont basées sur un réseau de neurones artificiels (ANN). Ils partagent une architecture commune. Les ANN sont composées d'unités connectées (neurones) disposées en couches. Les structures de connexion des unités cachées sont entièrement connectées ou connectées localement. La première couche recevant des données est la couche d'entrée, la dernière couche est appelée couche de sortie, enfin, les neurones entre les couches d'entrée et de sortie sont appelés unités cachées (figure 1). Chaque unité cachée stocke un ensemble de poids, qui seront ajustés durant l'apprentissage.

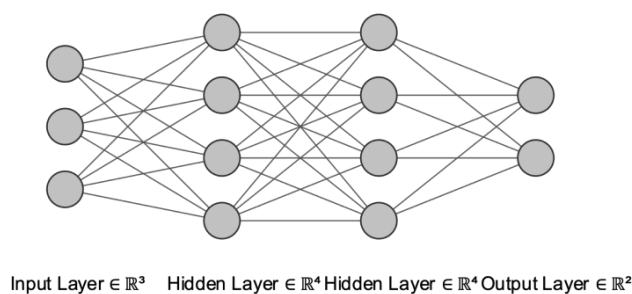


Figure 1 : Réseau neuronal avec une couche d'entrée, une couche de sortie et deux couches d'unités cachées.

Les ANN peuvent être supervisés, semi-supervisés ou non supervisés. Les approches de *deep learning* supervisé reposent sur une fonction  $y = f(x)$  qui associe l'entrée  $x$  à une sortie  $y$ . Les données étiquetées d'apprentissage  $x$  sont analysées par l'algorithme d'apprentissage supervisé et en est déduit une fonction qui peut être utilisée pour associer d'autres exemples. À l'inverse, les méthodes de *deep learning* non supervisées visent à apprendre des modèles et des propriétés inconnues d'intérêt sur la distribution de  $x$ ,  $x$  étant un ensemble de données non étiquetées. L'apprentissage des représentations des patients utilise principalement des méthodes supervisées [8,9], cependant, les techniques d'apprentissage non supervisées deviennent de plus en plus populaires [5] et permettent d'exploiter les données indépendamment de la présence d'étiquetage. Parmi les architectures de *deep learning* utilisées pour la représentation des patients, nous avons retenu 4 catégories principales. Tout d'abord, une approche supervisée commune utilisant un réseau neuronal récurrent (RNN) avec des unités de mémoire à court terme (LSTM) ou avec des unités récurrentes à déclenchement (GRU) [9–13], puis des architectures de réseau neuronal convolutif (CNN) [8]. Les architectures d'apprentissage non supervisées sont représentées par les auto-encodeurs (AE) [5,14] et l'apprentissage semi-supervisé par les transformateurs (TF) [15,16] (dans l'annexe 1 sont présenté les architectures des principaux réseaux de deep learning)

Les réseaux neuronaux convolutifs sont les réseaux neuronaux de type *feedforward* les plus populaires qui utilisent au moins dans une couche une opération mathématique différente (c'est-à-dire la convolution) à la place de la multiplication matricielle générale. Le modèle DeepR [8] utilise par exemple une couche d'intégration de mots (word2vec) dans une architecture CNN plus large pour prédire les réadmissions hospitalières non planifiées en prédisant les codes d'événements cliniques discrets. Une autre architecture d'apprentissage profond conventionnelle et sensible au temps sont les réseaux neuronaux récurrents. En ajoutant dans leurs architectures des connexions de rétroaction, les RNN ont permis de traiter des séquences temporelles. Les RNN s'appliquent particulièrement bien aux données ordonnées séquentiellement, comme les séries chronologiques ou le langage naturel. Le modèle DeepCare [9] utilise des cellules LSTM modifiées pour générer deux vecteurs, un diagnostic temporel et des codes d'intervention, qui sont ensuite concaténés pour obtenir une représentation du patient afin de prédire les prochains stades de la maladie et les réadmissions non planifiées. Avec un outil similaire, le modèle Doctor AI de Choi et al [12] utilisent une architecture RNN avec des unités GRU. Doctor AI utilise un *skip-gram embedding* comme schéma d'initialisation pour le RNN afin d'apprendre une représentation vectorielle efficace de l'état du patient à chaque horodatage. Le modèle Doctor AI utilise

des séquences d'événements et de temps comme entrée pour le réseau GRU. Ils ont été en mesure de prédire les catégories de diagnostics et de médicaments ainsi que la durée jusqu'à la prochaine visite. Ils ont également montré empiriquement que leur modèle était extensible à d'autres institutions médicales, même si les modèles avaient appris à partir de larges bases de données différentes. Cette généralisation de l'utilisation du modèle est cruciale, elle permet de l'utiliser pour des institutions médicales plus petites qui ne disposent pas de suffisamment de données sur les patients et donc d'entraîner un réseau de neurones sans sur-apprentissage (*overfitting*) (dans l'annexe 2 sont présentées les différentes limites de l'apprentissage d'un réseau de neurone).

D'autres auteurs ont incorporé des *revers-time attention mechanisms* dans les RNN afin d'obtenir de meilleures performances, comme Choi et al. avec le modèle RETAIN [13] plus performant que la plupart des autres modèles au moment de sa publication. Plus récemment Ma et al. ont proposé le modèle Health-ATM [17], une structure de réseau hybride composée à la fois de RNN et de CNN permettant d'extraire des structures exhaustives d'informations multidimensionnelles sur les patients.

Plusieurs nouvelles architectures ont récemment vu le jour, parmi lesquelles les auto-encodeurs et les transformateurs. Les auto-encodeurs sont *feedforward neural networks* à 3 couches (couche d'entrée, couche cachée et couche de décodage). L'entrée est codée dans une dimension inférieure, puis la sortie est reconstruite à partir de cette représentation. Un type courant d'AE est l'auto-encodeur de débruitage (DAE), où les entrées sont aléatoirement et partiellement corrompues, puis reconstruites. Les auto-encodeurs de débruitage empilés (SDAE) ont été introduits par Vincent et al [18], ils ont établi l'intérêt d'utiliser le critère de débruitage comme un moyen non supervisé pour guider l'apprentissage de la représentation de niveau supérieur. Miotto et al. [5] utilisent l'apprentissage profond non supervisé avec les SDAE [18] appliqué aux données du DPI pour apprendre une représentation du patient. Le modèle est utilisé secondairement pour prédire l'avenir des patients. Tout d'abord, Miotto et al. ont utilisé un apprentissage non supervisé des caractéristiques, ce qui a permis aux auteurs d'utiliser des données non étiquetées provenant d'une base de données de DPI à grande échelle. De plus, dans leur approche, les auteurs ont utilisé un réseau neuronal composé d'une superposition de trois couches d'auto-encodeurs de débruitage. Miotto et al. génèrent un vecteur patient à partir de données DPI brutes telles que des détails démographiques (c'est-à-dire l'âge, le sexe et la race) et des descripteurs cliniques communs disponibles dans un format structuré tels que les diagnostics (codes CIM-9), les médicaments, les procédures et les tests de laboratoire, ainsi que des notes cliniques en texte libre [5].

Enfin, l'architecture la plus récente utilisée pour apprendre à partir des données du DPI est celle des *Transformers*. Leur architecture est basée sur un ensemble d'encodeurs enchaînés. Chaque encodeur et décodeur utilise un *attention mechanism*. Les TF sont conçus comme des RNN pour traiter les séquences temporelles, néanmoins la séquence temporelle n'a pas besoin d'être traitée dans l'ordre contrairement aux RNN. Li et al. [15] ont proposé un modèle basé sur les *Transformers* pour les dossiers patient informatisés, appelé BEHRT.

Par ailleurs, Darabi et al avec le modèle TAPER [16] proposent un réseau de *Transformers* permettant de créer un plongement lexical (embedding) des données de DPI dans un vecteur de représentations unifiées. Pour atteindre cet objectif, Darabi et al, concatènent un code médical représentatif appris avec un modèle skip-gram et une représentation des notes médicales apprises avec le modèle BERT [19].

Les modèles de modélisations des données des patients utilisent différents types d'entrées et de processus de prétraitement des données. Leurs nuances sont résumées dans les tableaux 1, 2 et 3. Les modèles sont détaillés dans l'annexe 3.

	DeepR [8]	DeepCare [9]	Doctor AI [12]	RETAIN [13]
<b>Architecture type</b>	CNN	RNN (LSTM)	RNN (GRU)	RNN (GRU) Attention mechanism
<b>Word modeling &amp; Data pre-processing</b>	Word2Vec (unsupervised model relying on local collocation patterns, clinical motifs) Times of admission and discharge Level 3 ICD-10 codes Text (word embedding with word2vec)	Clinical concept word embedding with diagnoses and interventions (medications, procedures) Embedding of an admission vector (composed of diagnosis codes, intervention codes, admission method, elapsed time)	Skip-gram embeddings ICD-9 codes, medication codes, and procedure codes, grouped into level 3 ICD-9, and presented as sequences (event, time)	Code extractions ICD-9 code extraction GPI: Generic Product Identifier CPT: Current Procedural Terminology Medical grouper (codes aggregation) Codes sequence

Tableau 1 : Résumé du prétraitement des données et de type d'entrée des modèles de l'état de l'art

	Patient2Vec [20]	Darabi_2 [21]	Health-ATM [17]
<b>Architecture type</b>	RNN (GRU) Self-attention mechanism	RNN (GRU) + Transformer	RNN & CNN
<b>Word modeling &amp; Data pre-processing</b>	No word modeling Word2vec to construct a vector to represent each medical code Clinical visits are represented as the set of vectors for the codes associated with the visit Aggregation into universal vector + addition of static information (age, gender, previous hospitalization)	No word modeling Clinical time series & concepts embedding	Sequences of clinical events Event composed of a medical code and associated with a visiting time stamp

Tableau 2 : Résumé du prétraitement des données et de type d'entrée des modèles de l'état de l'art

	Sushil Model [14]	Deep Patient [5]	BEHRT [15]	TAPER [16]
<b>Architecture type</b>	SDAE	SDAE	Transformers	Transformers
<b>Word modeling &amp; Data pre-processing</b>	Tokenization of the dataset using the Ucto tokenizer	Pre-processing of free-text clinical notes with Open Biomedical Annotator	Pre-training with MLM (mask language model) 1) General demographic details (age, gender, ethnicity) 2) ICD-9 codes, medications, procedures, and lab tests, 3) Free-text clinical	1) Skip-gram: medical code representation 2) BERT model: medical note representation

Tableau 3 : Résumé du prétraitement des données et de type d'entrée des modèles de l'état de l'art

### 3 Matériel et méthodes

#### 3.1 Données et préparation des données

##### 3.1.1 Recueil des données

Les données ont été extraites de l'entrepôt de données de l'hôpital pour enfants de Necker. Pour chaque patient ont été extraits les codes CIM-10, les concepts de tests de laboratoire et de biologie, les concepts de notes médicales et les informations démographiques. Les concepts des notes médicales ont été extraits [22] précédemment à notre travail. Les données correspondent à une année de collecte, l'année de collecte 2014.

##### 3.1.2 Préparation des données : CIM-10 (CIM-10 vers UMLS CUI)

Les codes de la CIM-10 ont été mis en correspondance avec l'identifiant unique contrôlé du système de langage médical unifié (UMLS CUI) selon la procédure suivante : une première mise en correspondance a été effectuée entre les codes de la CIM-10 et l'UMLS CUI. Les codes non correspondants ont ensuite été mappés à la CIM-10-CM (modification clinique). Pour les codes non reconnus, nous avons procédé à un second mappage approximatif à partir des 3 premiers caractères de chaque code. Si le code complet était trouvé lors du premier mappage, il était conservé. Le deuxième mappage ne concerne que les codes précédents non reconnus, nous avons alors mappé uniquement les trois premiers caractères du code.

##### 3.1.3 Préparation des données : concepts des notes cliniques

L'extraction de concepts UMLS à partir de notes cliniques a déjà été réalisée sur l'entrepôt de données de l'hôpital pour enfants de Necker [23].

##### 3.1.4 Préparation des données : test de laboratoire (LOINC vers UMLS CUI)

En ce qui concerne les tests de laboratoire et les concepts biologiques, une première mise en correspondance a été réalisée entre les codes de Necker et une table de correspondance maison (ANABIO) avec les noms et codes des identificateurs logiques d'observation (LOINC). Ensuite, les codes LOINC ont été mis en correspondance avec l'UMLS CUI. Dans l'ensemble, nous avons pu obtenir 93,81 % de codes mappés.

##### 3.1.5 Préparation des données : données démographiques

Les informations démographiques (âge, sexe et statut de décès) ont été extraites du DPI.

##### 3.1.6 Vecteur patient

Enfin, nous avons utilisé comme données d'entrée un vecteur contenant le nombre de chaque UMLS CUI et les informations démographiques pour chaque patient.

	Sexe	Age	UMLS CUI 1	...	UMLS CUI n
Patient 1	<i>Sexe patient 1</i>	<i>Age patient 1</i>	<i>nombre d'UMLS CUI 1</i>	...	<i>nombre d'UMLS CUI n</i>
...	...	...	...	...	...
Patient n	<i>Sexe patient n</i>	<i>Age patient n</i>	<i>nombre d'UMLS CUI 1</i>	...	<i>nombre d'UMLS CUI n</i>

## 3.2 Implémentation

### 3.2.1 Architecture du modèle

Nous avons implémenté le modèle en Python avec comme *backend* TensorFlow et Keras en tant que bibliothèque de *deep learning*. Enfin, nous nous sommes appuyés sur l'implémentation de Sushil & al comme base de scripte [24].

Comme Deep Patient [5] nous avons utilisé des SDAE, entraînés de manière indépendante, couche par couche et basés sur le type d'architecture décrite par Vincent & al. [18].

#### 3.2.1.1 Auto-encodeurs (AE)

Un auto-codeur (AE) traditionnel est une structure d'apprentissage semi-supervisée composée d'un *encodeur* et d'un *décodeur* [18]. Il est donc constitué de 3 couches, la couche d'entrée, la couche cachée et la couche de sortie. (figure 2).

L'*encodeur* fait un mappage déterministe (*deterministic mapping*)  $f_{\theta}(x)$ ; type d'algorithme qui, étant donné une entrée particulière, produira toujours la même sortie. Ici l'encodeur transforme le vecteur d'entrée  $x$  en une représentation cachée  $y$  (couche cachée), avec  $x \in [0, 1]^d$  et  $y \in [0, 1]^{d'}$ , le processus d'encodage décrit par :

$$y = f_{\theta}(x) = s(Wx + b)$$

La fonction  $s(\cdot)$  est une transformation non linéaire appelée 'fonction d'activation'. Son jeu de paramètres est composé de  $W$  qui est la matrice de poids de l'encoder et  $b$ , un vecteur de décalage de dimension  $d'$  (vecteur de biais).

Le *Décodeur* utilise la représentation cachée résultante  $y$  et crée un mappage inverse vers un vecteur reconstruit  $z$ , avec  $z \in [0, 1]^d$ , par le processus décrit par :

$$z = g_{\theta'}(y) = s(W'y + b')$$

Son paramètre est  $\theta' = \{W', b'\}$ , où  $W'$  est la matrice de poids du décodeur et  $b'$  le vecteur de biais.

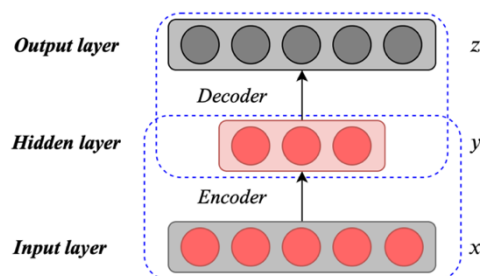


Figure 2 : Représentation schématique de l'architecture d'un auto-encodeur (AE)

L'auto-codeur actuel est composé d'un encodeur affine et sigmoïde et d'un décodeur affine

et sigmoïde avec une fonction objectif (*loss function*) de type entropie croisée (*cross-entropy*). La contrainte d'utiliser des poids liés (*tied weights*) entre  $W$  et  $W'$  est utilisée et définie par  $W'$  comme  $W' = W^T$ . Ainsi les poids de l'encodeur sont transposés au décodeur permettant de réduire le nombre de paramètres du modèle [18].

Le jeu de paramètre de l'auto-encodeur est optimisé pour réduire l'erreur de reconstruction selon la fonction objectif :

$$\operatorname{argmin}_{\theta, \theta'} \mathbb{E}_{q^0(X)} [L(X, Z(X))]$$

### 3.2.1.2 Auto-encodeur de débruitage (DAE)

L'auto-codeur de débruitage est une variante de l'auto-codeur de base décrit ci-dessus. Lors de l'apprentissage du modèle, l'algorithme recherche les paramètres qui minimisent la différence entre  $x$  et  $z$  (i.e., l'erreur de reconstruction  $L_H(x, z)$ ).

Le signal d'entrée  $x$  est corrompu en  $\tilde{x}$  au moyen d'un mappage stochastique  $\tilde{x} \sim q_D(\tilde{x}|x)$ . Le mappage stochastique est fonction de mappage aléatoire permettant de créer du bruit dans le signal d'entrée. Le signal d'entrée corrompue  $\tilde{x}$  est ensuite mappé dans un auto-encodeur de base :

$$y = f_{\theta}(\tilde{x}) = s(W\tilde{x} + b)$$

A partir duquel est reconstruit  $z$ .

$$z = g_{\theta'}(y)$$

L'objectif est de reconstruire les données d'entrée. Au cours du processus, la couche cachée capture les caractéristiques utiles de la couche d'entrée. L'algorithme de descente de gradient stochastique (*stochastic gradient descent algorithm*) est utilisé pour la mise à jour des poids dans l'objectif de minimiser l'erreur de reconstruction d'entropie croisée  $L_H$ .

L'architecture générale de l'auto-encodeur de débruitage est présentée dans la figure 3.

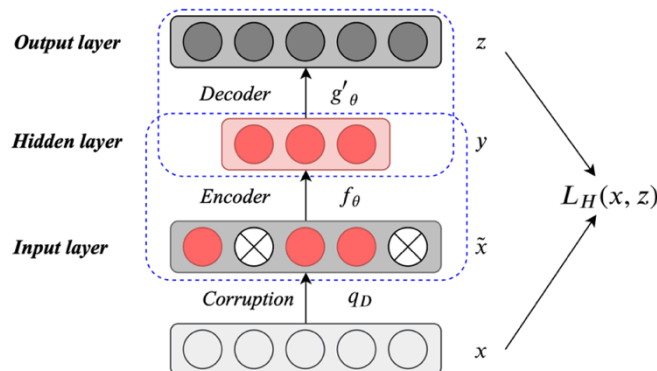


Figure 3 : Représentation schématique de l'architecture d'un auto-encodeur de débruitage (DAE)

### 3.2.1.3 Auto-encodeur de débruitage superposés (SDAE)

Dans un SDAE, l'entrée initiale est utilisée pour obtenir une première représentation supérieure (figure 4, DAE1), ensuite le vecteur de caractéristique (couche cachée) est utilisé comme donnée d'entrée de l'auto-encodeur de débruitage suivant (figure 4, DAE2). Le processus peut être répété autant de fois qu'il existe de DAE superposés dans le modèle. Chaque fois, une représentation de niveau supérieur est extraite de la couche cachée. Finalement, un classifieur est ajouté au bout du réseau et de paramètre  $\theta = \{W, b\}$ . Ces paramètres sont simultanément ajustés (*fine-tuned*) grâce à une descente de gradient stochastique (figure 4).

Finalement, la représentation de niveau la plus supérieure est utilisée comme donnée d'entrée pour des sous-tâches indépendantes. Cette représentation est réutilisable pour différentes sous-tâches, dans cette étude nous procédons à une tâche de classification.

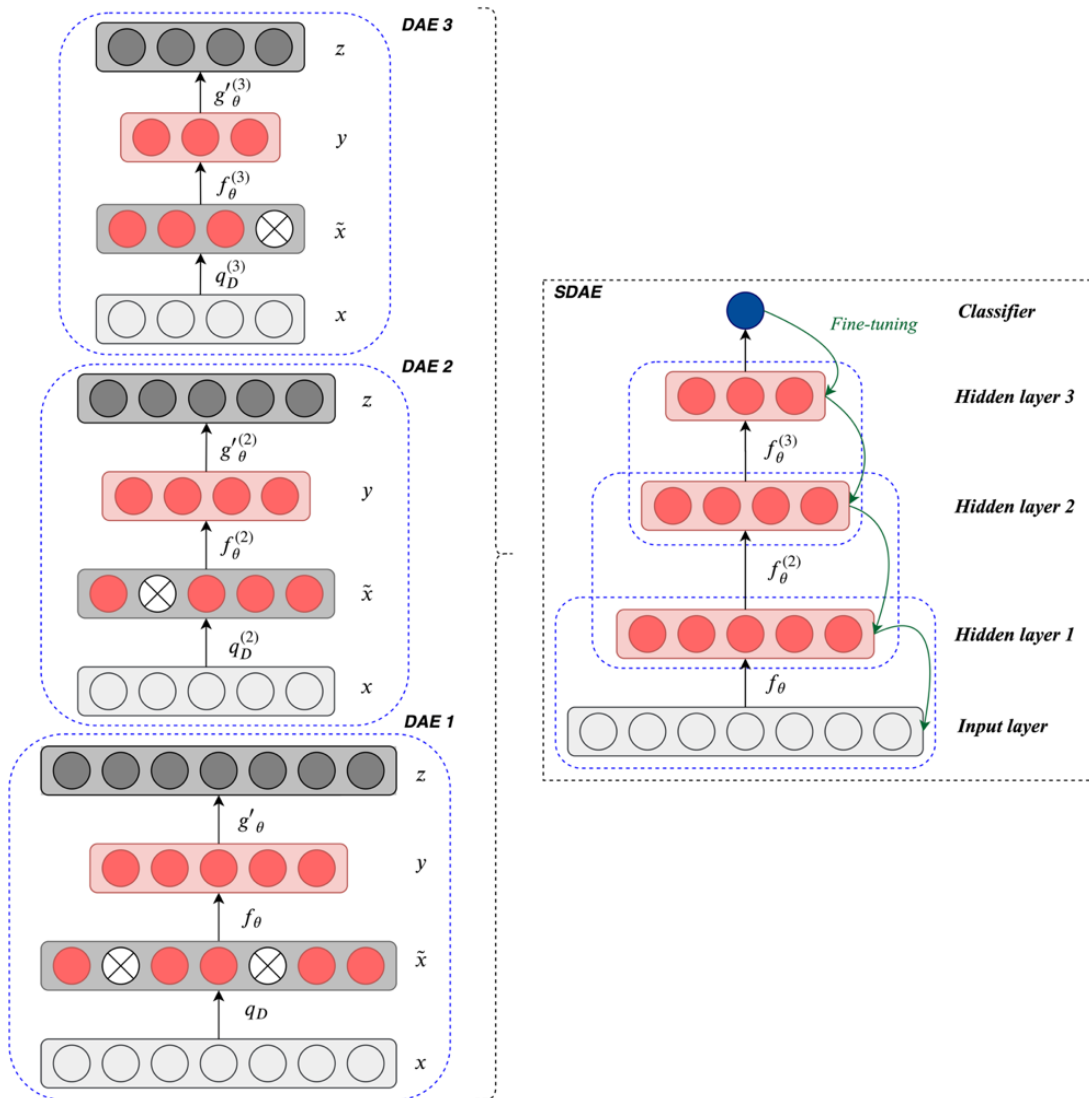


Figure 4 : Représentation schématique de l'architecture d'un auto-encodeur de débruitage superposé (SDAE)



### 3.2.2 Représentation réutilisable du patient : entraînement et hyperparamètres

Le réseau a d'abord été entraîné selon le paramétrage du réseau de Deep Patient [5], chaque auto-codeur était composé de couches entièrement connectées entraînées de manière avide (*greedy training*) [25,26] avec des poids liés (*tied weights*). Nous avons utilisé une fonction d'activation sigmoïde pour l'encodeur et une fonction d'activation linéaire pour le décodeur. Chaque couche dense (encodeur et décodeur) était composée de 500 unités cachées. Le bruit d'entrée a été inséré avec une couche de décrochage (*dropout*), masquant 5% des données d'entrée. Nous avons utilisé 20 époques par auto-encodeur et utilisé une méthode d'arrêt précoce de l'apprentissage pour éviter le sur-apprentissage (*overfitting*). La méthode du descente de gradient a été utilisée pour mettre à jour les paramètres  $\theta = \{W, b\}$ .

### 3.2.3 Sous-tâches (ré-hospitalization) : entraînement et hyperparamètres

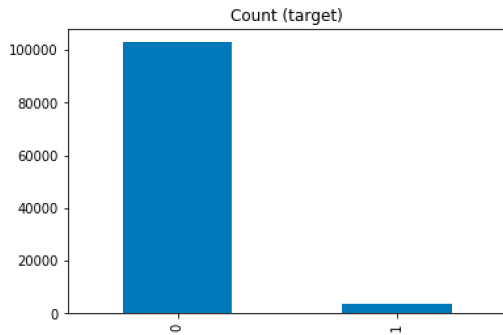
Pour tester notre représentation du patient, un sous-ensemble de données sur les réhospitalisations a été créé. Nous avons utilisé le tableau des documents et sélectionné uniquement le document de sortie de l'hôpital. Nous avons ensuite calculé le delta de temps entre le premier document de sortie et le suivant. Si le delta était inférieur à 30 jours, l'hospitalisation était étiquetée comme vraie, s'il y avait une seule hospitalisation ou si le delta était supérieur à 30 jours, l'hospitalisation était étiquetée comme fausse.

Nous avons ensuite testé la capacité de la représentation profonde des patients à déterminer si les patients étaient réhospitalisés en moins de 30 jours. Un classificateur à couche dense a été ajouté au SDAE ; nous avons utilisé les mêmes paramètres d'ajustement que pour construire la représentation et une fonction d'entropie croisée catégorielle (*cross entropy loss function*). Pour déterminer la précision et le F-score, un seuil a été fixé pour distinguer les prédictions positives des négatives. Ce seuil a été fixé à 0,6 et déterminé à partir de l'ensemble de validation. Cette valeur a été choisie afin d'optimiser en réduisant les prédictions faussement positives.

### 3.2.4 Modèle de références (baselines)

Nous avons utilisé une forêt aléatoire (*random forest*) et une machine à vecteurs de support (SVM) comme modèle de référence d'apprentissage automatique. De plus, nous avons implémenté un auto-encodeur profond (DAE) comme modèle de référence d'apprentissage profond. Le DAE était composé de deux couches d'encodeurs et de deux couches de décodeurs, la couche d'entrée avait 200 unités cachées, le premier décodeur 100 et le dernier 50. Nous avons utilisé une fonction d'activation *tanh* pour la couche d'entrée. Les couches encodeurs ont utilisé la fonction d'activation *relu* et les couches décodeurs la fonction d'activation *tanh*. Enfin, nous avons utilisé *adadelta* comme optimiseur, *mse* comme fonction objectif (*loss function*), un nombre d'exemples (batch size) de 256 et utilisé 100 époques. Nous avons testé les modèles de références sur des ensembles de données de différentes tailles (1000 patients, 10000 patients et toutes les données). Nous avons également testé sur un échantillon réduit (*down sample*) de l'ensemble d'entraînement pour éviter le paradoxe de précision (figure 5) [27]. Le paradoxe de précision se rencontre dans les problèmes de classification où une classe est largement supérieure à l'autre. Ainsi l'algorithme de classification semble très performant car il prédit toujours correctement la classe majoritaire. L'échantillon réduit permet d'obtenir des échantillons égaux entre les deux classes étudiées et de s'affranchir du paradoxe.

Class 0: 102889  
 Class 1: 3787  
 Proportion: 27.17 : 1



Random under-sampling:  
 1 3787  
 0 3787  
 Name: target, dtype: int64

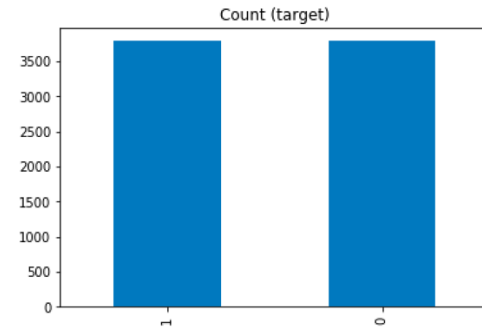


Figure 5 : Échantillon réduit

## 4 Résultats

### 4.1 Description de la population

#### 4.1.1 Données complètes

La population étudiée présentait une répartition homogène des sexes, l'âge moyen était de 23,73 ans (max = 110 et min 1). (tableau 4&5, figure 6).

	Population	Percentage
<b>Total</b>	103729	100%
<b>Female</b>	53433	50.33%*
<b>Male</b>	52719	49.66%*

\* without missing values

	Mean age (year)	Max age (year)	Min age (year)
<b>Patients total</b>	23.73	110	1
<b>Female</b>	25.44	110	1
<b>Male</b>	22.20	106	1

Tableau 4 & 5 : Démographie et âge de la population

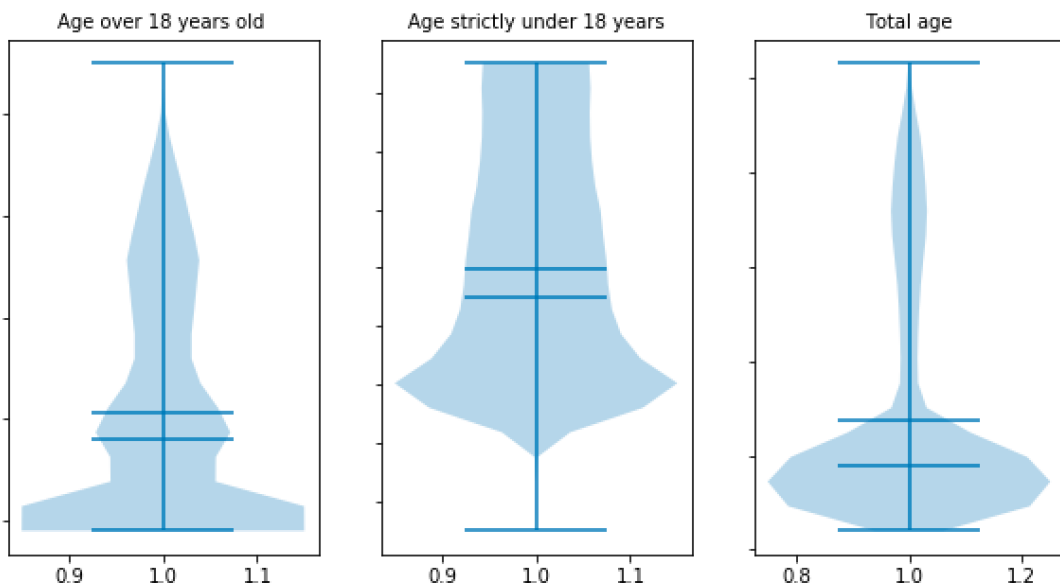


Figure 6 : Graphique en violon de la distribution des âges

Nous avons observé un âge moyen de 23 ans arrondi qui s'explique par le fait que l'hôpital Necker accueille peu de patients adultes. L'étude démographique montre une majorité de patients de moins de 18 ans (59898 patients) et 47065 patients de plus de 18 ans, respectivement 56% et 44%. Nous avons choisi pour cette étude de conserver la base de données totale et de tester uniquement le SDAE sur la population pédiatrique.

#### 4.1.2 Données pédiatriques

Nous avons suivi le même traitement des données et le même mappage pour la population de patients pédiatriques (tableau 6 & 7 et figure 7). Nous avons choisi de représenter la population intersexe dans l'ensemble des données pédiatriques car elle était plus importante que dans la population totale.

	Population	Percentage	Mean age (year)	Median (year)	Max age (year)	Min age (year)
<b>Total</b>	59695	100%	10.39	10	18	1
<b>Female</b>	26708	44.74%	10,40	10	18	1
<b>Male</b>	32985	55.26%	10.33	10	18	1
<b>Intersex</b>	2	0.0033%	7.7	6.5	14	5

Tableau 6 & 7 : Démographie et âge de la population pédiatrique

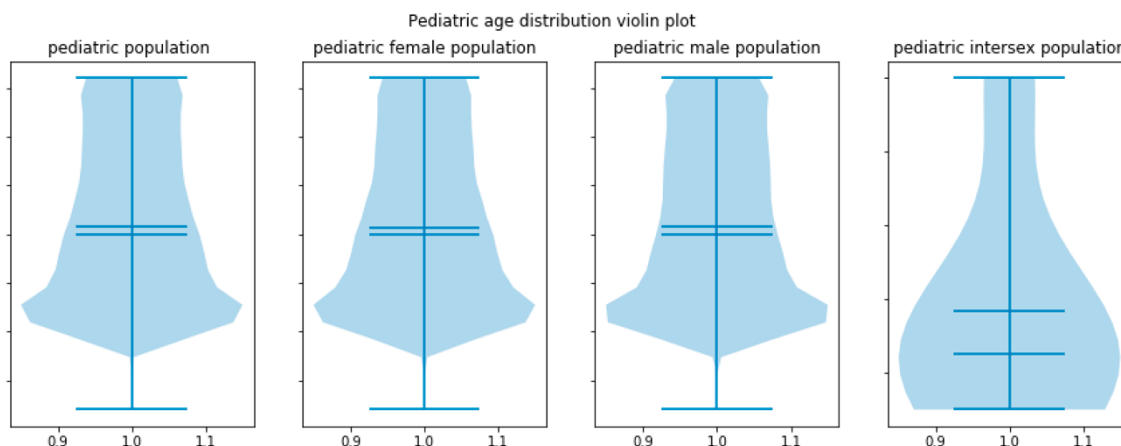


Figure 7 : Graphique en violon de la distribution de l'âge par sexe pour la population pédiatrique

## 4.2 Préparation des données et mappage

### 4.2.1 Données complètes

Cette première phase de mappage nous a permis d'obtenir 91,63% de correspondance exacte. La deuxième phase de mappage nous a permis de récupérer 6,57 % de code sans correspondance après la première étape. Dans l'ensemble, 98,2 % des codes CIM-10 ont été mis en correspondance avec les UMLS CUI (tableau 8).

Number of	total of total CUI mapped	6069486
	unique CUI	10645
	biology unique CUI	935
	unique CIM CUI	5366
	concept text CUI	4472

Tableau 8 : répartition des codes UMLS CUI

Parmi les CUI, nous récupérons 935 CUI de biologie uniques, 5366 CUI de CIM uniques et 4472 CUI de concept de texte unique. Le vecteur final est composé de 10645 CUI avec 8% de CUI de biologie, 50% de CUI de CIM et enfin 40% de CUI de concept de texte. Il existe un léger chevauchement entre les CUI des différents tableaux. Ainsi, nous avons pu trouver des CUI identiques entre les tableaux de biologie, les concepts CIM et les textes (128 CUI se chevauchent). Enfin, il y avait une moyenne de 58 CUI par patient.

La base de données s'est avérée déséquilibrée (tableau 9) avec 103729 patients, 4211 réhospitalisations (3,94%) après 30 jours et 3787 réhospitalisations (3,55%) avant 30 jours (total de 7998 réhospitalisations).

Number total of rehospitalization	7998
Number of rehospitalization in less than 30 days	3787

Tableau 9 : Réhospitalisations

La rareté des événements de réhospitalisation à moins de 30 jours est décrite dans la figure 8. On peut voir la proximité graphique entre les patients non réhospitalisés et les patients réhospitalisés à moins de 30 jours qui sont donc difficiles à classer avec précision à partir d'un modèle.

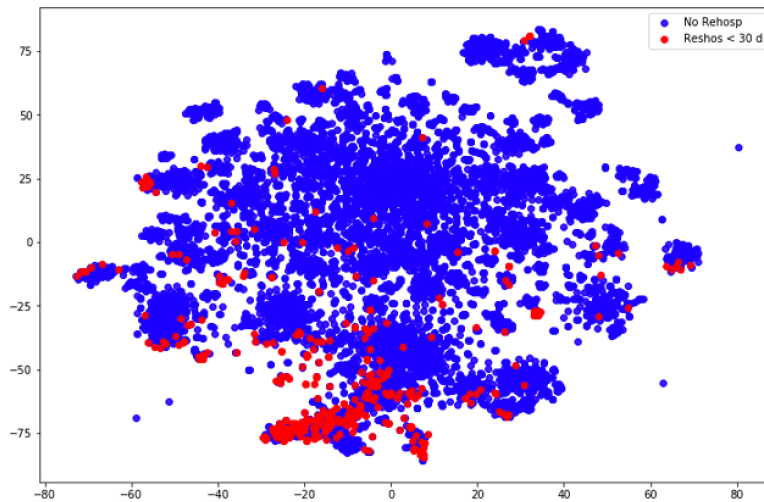


Figure 8 : Représentation graphique des données

#### 4.2.2 Données pédiatriques

L'ensemble de données pédiatriques présentait les mêmes caractéristiques que l'ensemble des données, avec 10644 CUI, mais la moyenne des CUI par patient était plus élevée (moyenne de 82,29 par patient).

#### 4.3 Phase d'apprentissage de l'auto-encodeur de débruitage superposés

Pendant la phase d'apprentissage, nous avons testé plusieurs configurations pour SDAE. Nous avons sélectionné la configuration avec la meilleure performance. Elle était composée de 500 unités cachées au sein de chaque auto-encodeur, dropout : 0.05, fonction d'activation sigmoïde pour l'encodeur, fonction d'activation linéaire pour le décodeur, 20 époques, nombre d'exemples : 512. Enfin nous avons utilisé mse pour fonction d'objectif et rmsprop comme optimiseur. Nous avons utilisé un générateur de nombre d'exemples (*batch generator*) pour charger les données dans le SDAE.

## 4.4 Performances d'auto-encodeur de débruitages superposés

### 4.4.1 Base de données complète

Les résultats des modèles de références sont décrits dans le tableau 10, les résultats des SDAE dans le tableau 11.

	Sample variation	Precision (False / True)	Recall (False / True)	F-score (False / True)	Accuracy
Random forest 1000	1000	0.97 / 0.33	0.99 / 0.14	0.98 / 0.20	0.96
	10000	0.97 / 0.60	1.00 / 0.12	0.98 / 0.20	0.96
	All data	0.98 / 0.56	0.99 / 0.31	0.98 / 0.40	0.96
	Down-sample	1.00 / 0.29	0.91 / 1.00	0.95 / 0.45	0.92
Support vector machine (SVM)	1000	0.97 / 0.20	0.98 / 0.14	0.97 / 0.17	0.93
	10000	0.98 / 0.44	0.98 / 0.46	0.98 / 0.45	0.95
	All data	0.73 / 0.56	0.74 / 0.55	0.74 / 0.55	0.667
	Down-sample	1.00 / 0.29	0.91 / 1.00	0.95 / 0.45	0.92
Deep AE	1000	0.98 / 0.40	0.99 / 0.33	0.99 / 0.36	0.97
	10000	0.97 / 0.42	0.99 / 0.17	0.98 / 0.24	0.96
	All data	0.97 / 0.39	0.99 / 0.12	0.98 / 0.19	0.96
	Down-sample	0.81 / 0.90	0.90 / 0.80	0.85 / 0.85	0.85

Tableau 10 : Résultats des modèles de références pour la base de donnée totale

	Sample variation	Precision (False / True)	Recall (False / True)	F-score (False / True)	Accuracy
SDAE	1000	1.00 / 0.45	0.94 / 1.00	0.97 / 0.62	0.94
	10000	1.00 / 0.41	0.95 / 0.91	0.97 / 0.57	0.95
	All data	0.99 / 0.76	0.99 / 0.79	0.99 / 0.78	0.98
	Down-sample	0.50 / 0.59	0.90 / 0.14	0.64 / 0.23	0.51

Tableau 11 : Résultat du SDEA pour la base de donnée totale

### 4.4.2 Données pédiatriques

Pour les données pédiatriques nous avons effectué l'analyse uniquement sur l'échantillon complet (tableau 12).

	Precision (False / True)	Recall (False / True)	F-score (False / True)	Accuracy
SDAE	0.99 / 0.80	0.99 / 0.86	0.99 / 0.86	0.90

Tableau 12 : Résultat du SDEA pour les données pédiatriques

## 5 Discussion

Ce travail présente une représentation profonde du dossier patient (*deep representation of patient*) en utilisant uniquement les données brutes du DPI. Cette représentation peut être utilisée pour de multiples sous-tâches comme des tâches de prédiction ou de clustering.

Nos résultats montrent une meilleure prédiction pour le SDAE dans toutes les configurations testées que pour les modèles de références (forêt aléatoire et SVM). Cependant le réseau semble être moins efficace que dans d'autres études similaires, en particulier pour Deep Patient [5]. La forêt aléatoire et le SVM ont montré de faibles performances. En raison du déséquilibre de l'ensemble de données, le classificateur prédit toujours la classe la plus courante c'est-à-dire l'absence de réhospitalisation à moins de 30 jours, sans effectuer aucune analyse des caractéristiques, la précision est donc élevée mais trompeuse. c'est le paradoxe de précision régulièrement rencontré en machine learning [28]. Les variations de l'échantillon n'ont pas modifié la précision, sauf pour le SVM sur l'ensemble des données qui a atteint une précision de 0,66 avec une précision sur la réhospitalisation de 0,56. Le sous-échantillonnage de l'échantillon d'entraînement n'a pas amélioré les résultats des modèles de références, à l'exception de l'auto-encodeur profond qui a obtenu de bons résultats sur le sous-échantillon pour la tâche de prédiction.

L'augmentation des échantillons permet une augmentation de la performance pour les modèles de référence et les SDAE. C'est généralement un effet constaté dans la plupart des problèmes de deep learning à variance élevée, plus le dataset est important plus l'algorithme sera performant [29,30]. Néanmoins, dans cette étude, le SDAE présente une meilleure précision pour la prédiction sur l'échantillon pédiatrique (0.80) que pour l'échantillon total (0.76) alors que celui-ci est environ deux fois plus petit (59695 patient pour l'échantillon pédiatrique et 103729 patient pour l'échantillon total). L'hypothèse la plus probable pour expliquer cette différence de performance ne semble pas être dans la taille de l'échantillon car notre dataset présente une variance élevée (10645 CUI pour le vecteur de caractéristique) mais car la base pédiatrique présentait plus de CUI moyen par patient (82,29) que la base total (58) permettant ainsi de classer avec plus de précision les patients.

Le vecteur de caractéristiques est plus grand dans notre étude (10645) comparé à 500 pour Deep Patient [5], cela pourrait être la raison principale des moins bonnes performances. La représentation semble être trop éparsée pour que le SDAE puisse construire une bonne représentation.

Un encastrement des concepts (*concept embeddings*) CUI aurait pu être effectué, Cui2vec a été récemment développé par Beam et al. [31] et est un moyen plus efficace de représenter des données clairsemées et de dimension élevée. Une représentation des CUIs de dimension inférieure pourraient permettre à l'algorithme d'apprentissage profond d'être plus efficace. D'autres encastresments (*embeddings*) auraient pu être ajoutés pour améliorer la représentation, en particulier un encastrement temporel des CUIs.

Nous n'avons réalisé qu'un ajustement (*fine-tuning*) des paramètres au début de l'implémentation du modèle. Il est certain qu'un ajustement des paramètres sur les modèles pré-entraînés aurait amélioré les performances. De même, un modèle de référence de modélisation de texte à la pointe de la recherche pourrait être utilisé à la place de l'extraction des concepts textuels utilisés dans cette étude. Une autre possibilité à explorer serait d'ajouter plusieurs autres sources de données au modèle. Le modèle pourrait inclure par exemple des données d'imagerie, des données biométriques, des données plus complexes telles que des données d'électrocardiogramme, des données sur les traitements (médicamenteux, de radiothérapie, etc) ainsi que les données de mouvements du patient dans l'hôpital. De plus, même si les SDAE sont des moyens performants pour construire une représentation inférieure, d'autres algorithmes d'apprentissage profond pourraient être

performants pour ce type de tâches, en particulier en utilisant de nouvelles architectures telles que les transformers.

Pour de futures recherches, il serait intéressant de tester ce modèle sur différentes sous-tâches, en particulier le clustering pour la prédiction diagnostic, car il pourrait permettre un diagnostic plus précoce de maladies rares. Les améliorations possibles sont considérables, multiples et une application uniquement aux maladies rares serait pertinente.

## Bibliographie

1. Bernard A. Clinical prediction models: a fashion or a necessity in medicine? *J Thorac Dis.* oct 2017;9(10):3456-7.
2. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating* [Internet]. Cham: Springer International Publishing; 2019 [cité 29 juin 2020]. (Statistics for Biology and Health). Disponible sur: <http://link.springer.com/10.1007/978-3-030-16399-0>
3. The UK Strategy for Rare Diseases. [Internet]. Disponible sur: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/260562/UK\\_Strategy\\_for\\_Rare\\_Diseases.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pdf)
4. Plan national français pour les maladies rares 2018-2022 [Internet]. Disponible sur: [https://solidarites-sante.gouv.fr/IMG/pdf/pnmr\\_3\\_v25-09pdf.pdf](https://solidarites-sante.gouv.fr/IMG/pdf/pnmr_3_v25-09pdf.pdf)
5. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci Rep.* mai 2016;6(1):26094.
6. Dechter R. Learning While Searching in Constraint-satisfaction-problems. In: *Proceedings of the Fifth AAAI National Conference on Artificial Intelligence* [Internet]. AAAI Press; 1986. p. 178-83. (AAAI'86). Disponible sur: <http://dl.acm.org/citation.cfm?id=2887770.2887799>
7. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE J Biomed Health Inform.* sept 2018;22(5):1589-604.
8. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. Deepr: A Convolutional Net for Medical Records. *IEEE J Biomed Health Inform.* janv 2017;21(1):22-30.
9. Pham T, Tran T, Phung D, Venkatesh S. DeepCare: A Deep Dynamic Memory Model for Predictive Medicine. Bailey J, Khan L, Washio T, Dobbie G, Huang JZ, Wang R, éditeurs. *Adv Knowl Discov Data Min.* 2016;9652:30-41.
10. Jagannatha A, Yu H. Structured prediction models for RNN based sequence labeling in clinical text. *ArXiv160800612 Cs* [Internet]. août 2016 [cité 3 déc 2019]; Disponible sur: <http://arxiv.org/abs/1608.00612>
11. Esteban C, Schmidt D, Krompass D, Tresp V. Predicting Sequences of Clinical Events by Using a Personalized Temporal Latent Embedding Model. In: *2015 International Conference on Healthcare Informatics* [Internet]. Dallas, TX, USA: IEEE; 2015 [cité 3 déc 2019]. p. 130-9. Disponible sur: <http://ieeexplore.ieee.org/document/7349684/>
12. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *ArXiv151105942 Cs* [Internet]. sept 2016 [cité 2 déc 2019]; Disponible sur: <http://arxiv.org/abs/1511.05942>
13. Choi E, Bahadori MT, Kulas JA, Schuetz A, Stewart WF, Sun J. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. *ArXiv160805745 Cs* [Internet]. 26 févr 2017 [cité 6 déc 2019]; Disponible sur: <http://arxiv.org/abs/1608.05745>
14. Sushil M, Šuster S, Luyckx K, Daelemans W. Patient representation learning and

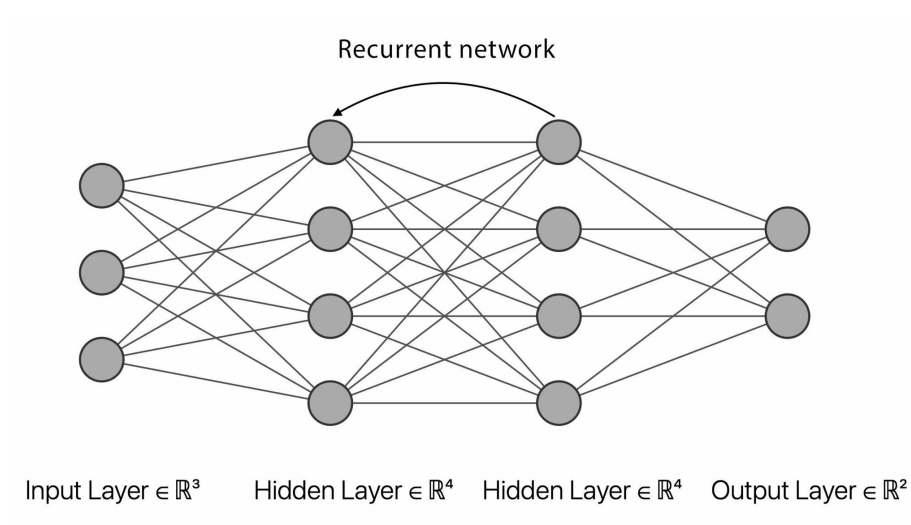


- interpretable evaluation using clinical notes. *J Biomed Inform.* août 2018;84:103-13.
15. Li Y, Rao S, Solares JRA, Hassaine A, Canoy D, Zhu Y, et al. BEHRT: Transformer for Electronic Health Records. *ArXiv190709538 Cs Stat [Internet]*. 22 juill 2019 [cité 6 déc 2019]; Disponible sur: <http://arxiv.org/abs/1907.09538>
  16. Darabi S, Kachuee M, Fazeli S, Sarrafzadeh M. TAPER: Time-Aware Patient EHR Representation. *ArXiv190803971 Cs Stat [Internet]*. août 2019 [cité 4 déc 2019]; Disponible sur: <http://arxiv.org/abs/1908.03971>
  17. Ma T, Xiao C, Wang F. Health-ATM: A Deep Architecture for Multifaceted Patient Health Record Representation and Risk Prediction. *SIAM Int Conf Data Min.* sept 2019;
  18. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J Mach Learn Res.* 2010;11:3371-408.
  19. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv181004805 Cs [Internet]*. 24 mai 2019 [cité 16 déc 2019]; Disponible sur: <http://arxiv.org/abs/1810.04805>
  20. Zhang J, Kowsari K, Harrison JH, Lobo JM, Barnes LE. Patient2Vec: A Personalized Interpretable Deep Representation of the Longitudinal Electronic Health Record. *IEEE Access.* 2018;6:65333-46.
  21. Darabi S, Kachuee M, Sarrafzadeh M. Unsupervised Representation for EHR Signals and Codes as Patient Status Vector. *ArXiv191001803 Cs Stat [Internet]*. 4 oct 2019 [cité 10 déc 2019]; Disponible sur: <http://arxiv.org/abs/1910.01803>
  22. Garcelon N, Neuraz A, Salomon R, Faour H, Benoit V, Delapalme A, et al. A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse. *J Biomed Inform.* avr 2018;80:52-63.
  23. Garcelon N, Neuraz A, Benoit V, Salomon R, Burgun A. Improving a full-text search engine: the importance of negation detection and family history context to identify cases in a biomedical data warehouse. *J Am Med Inform Assoc.* 20 oct 2016;ocw144.
  24. Sushil M. SDAE implementation [Internet]. 2019. Disponible sur: <https://github.com/MadhumitaSushil/SDAE>
  25. Bengio Y, Lamblin P, Popovici D, Larochelle H, Montreal U. Greedy layer-wise training of deep networks. Vol. 19, *Advances in Neural Information Processing Systems.* 2007.
  26. Hinton GE, Osindero S, Teh Y-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* juill 2006;18(7):1527-54.
  27. Barros TM, SouzaNeto PA, Silva I, Guedes LA. Predictive Models for Imbalanced Data: A School Dropout Perspective. *Educ Sci.* 15 nov 2019;9(4):275.
  28. Forner P, Müller H, Paredes R, Rosso P, Stein B, éditeurs. *Information Access Evaluation. Multilinguality, Multimodality, and Visualization: 4th International Conference of the CLEF Initiative, CLEF 2013, Valencia, Spain, September 23-26, 2013. Proceedings.* 1st ed. 2013. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer; 2013. 1 p. (Information Systems and Applications, incl. Internet/Web, and HCI).
  29. Banko M, Brill E. Scaling to Very Very Large Corpora for Natural Language Disambiguation. In: *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics [Internet]*. USA: Association for Computational Linguistics; 2001. p. 26-33. (ACL '01). Disponible sur: <https://doi.org/10.3115/1073012.1073017>
  30. Halevy A, Norvig P, Pereira F. The Unreasonable Effectiveness of Data. *IEEE Intell Syst.* mars 2009;24(2):8-12.

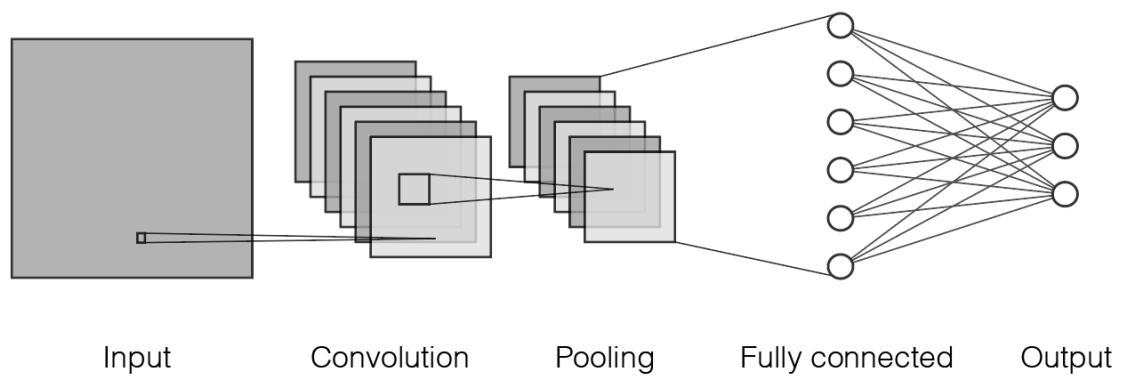
31. Beam AL, Kompa B, Schmaltz A, Fried I, Weber G, Palmer NP, et al. Clinical Concept Embeddings Learned from Massive Sources of Multimodal Medical Data. ArXiv180401486 Cs Stat [Internet]. 19 août 2019 [cité 17 janv 2020]; Disponible sur: <http://arxiv.org/abs/1804.01486>

## Annexe 1 : Présentation des architectures principal des réseaux de neurones

### Réseau neuronaux récurrent



### Réseau neuronaux convolutionnel



**Annexe 2** : Présentation des limites de l'apprentissage des réseaux de neurones (notion *overfitting* et *underfitting*)

### 1) Overfitting (surapprentissage)

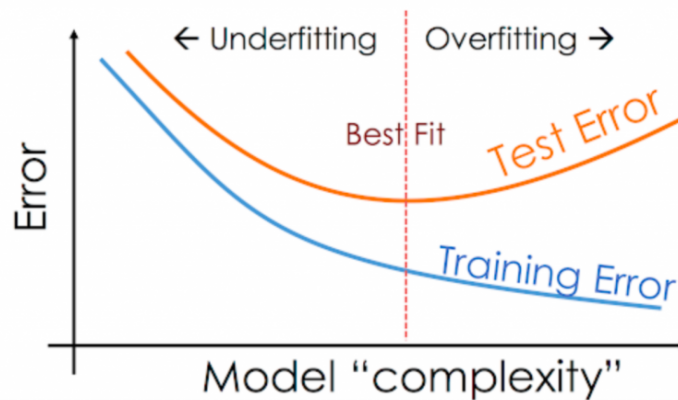
L'*overfitting* fait référence à un modèle qui modélise trop bien les données d'entraînement. C'est-à-dire lorsqu'un modèle apprend les détails et le bruit des données d'apprentissage qui sont alors appris comme concept par le modèle. Un modèle qui *overfit* est un modèle qui ne pourra pas généraliser à d'autres données.

### 2) Underfitting (sous-apprentissage)

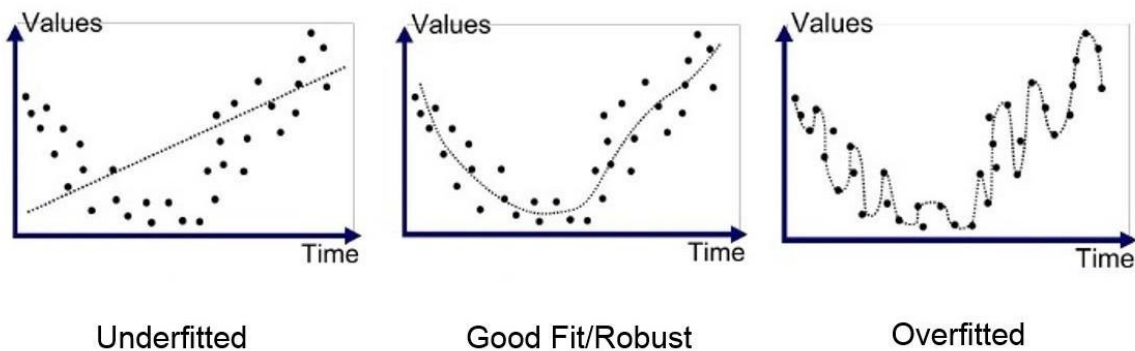
L'*underfitting* fait référence à un modèle qui ne peut ni modéliser les données d'apprentissage ni généraliser à de nouvelles données. Les raisons peuvent être multiples, généralement il signifie que le modèle choisi n'est pas adapté aux données.

### 3) Good fit (bon apprentissage)

L'objectif est d'avoir un bon ajustement entre ses données et son modèle d'apprentissage. C'est-à-dire ne pas être en *overfitting*, ni en *underfitting* mais entre les deux. Cette question est au centre des problématiques de la recherche en deep learning.



Représentation graphique durant de l'apprentissage



Représentation graphique de l'ajustement du modèle par rapport aux données

### Annexe 3 : Présentation détaillés de l'état de l'art

Tableau 1: Deep learning methods to learn patient representation using stacked denoising encoders

	Deep Patient [1]		Sushil's Model [10]
<b>Date / Author / Jr</b>	Mai 2016 / Miotto & al. / Scientific Reports		June 2018 / Sushil & al. / Elsevier
<b>Purpose</b>	This method captures hierarchical regularities and dependencies in the data to create a compact, general-purpose set of patient features that can be effectively used in predictive clinical applications.		Learning patient representations only from clinical notes for an arbitrary task, while understanding model performance.
<b>Method:</b> architecture (auto-encoder / classifier)	Multi-layer neural network. At every level: more abstract representation (higher level representation) with non-linear operation. Stack of denoising autoencoders (SDA), independently trained layer by layer. Trained on noisy version of the original data with the masking noise algorithm in order to prevent overfitting à 3-layer stack of denoising autoencoders were used in the initial unsupervised training stage. Random forest classifier was subsequently used in the last step of disease prediction. Hyperparameter tuning for autoencoders was performed using a separate validation dataset (tuned by looking at the supervised classification performance).		Stacked denoising autoencoder: trained as an independent The number of layers is determined through a random search based on the results for primary diagnostic category prediction using a perceptron. Classifiers are trained for different tasks using R as the input features. <b>Paragraph vector (Doc2vec):</b> training of 300-dimensional patient vectors (randomized hyperparameter search to determine the settings). <b>Feature extraction</b> (make dense representations interpretable)
<b>Hyper-parameters</b>	3 layers stack of denoising autoencoders (more layers have been reported, with no obvious improvement). 500 hidden units per layer / 5% noise corruption factor (used zero-out masking noise) / Mini-batch stochastic gradient descent / Sigmoid activation function / Cross-entropy loss function / Noise corruption factor $\nu = 5\%$ / 100 random forest trees.		Encoding layers: sigmoid activation function for the, Decoding real values: linear activation function. Pretraining phase: each layer of the SDAE is trained to minimize the mean squared reconstruction error using the RMSProp optimizer. GridSearch to determine (for stack denoising autoencoder) number of layers (1) / dimensionality (800-300) / dropout proportion (0.05-0.4) / learning rate (0.001). Hyperparameters for the feedforward neural network are determined for each task.
<b>Data set:</b> Data description Type of data Pretreatments	EHR data from Mont Sinai Filtered by thresholding on ICD9 codes (frequencies). Patient features include demographics, structured clinical descriptors (ICD9), medication, procedures, lab tests, and text mining (NLP) features (mainly topics from topic modeling). Patient disease labels were obtained by simple disease categorization from ICD9 codes, representing 78 diseases after filtering.	Data in year 2014 were retained as testing, and data before that as training. Final training data: 704,587 patients, and 41,072 features : Training: 704,587 patients (unsupervised training). / Validation: 5,000 patients (supervised evaluation) / Testing: 76,214 patients (supervised evaluation).	MIMIC-III database 24,650 patients for training 3081 patients each for validation and testing Patients are represented with a concatenation of all the notes associated with them (excluding discharge reports), then the dataset is tokenized using the Ucto tokenizer and lowercase it. Was used BoW feature finding a vocabulary size of 71,001 for the set, and the BoCUI feature finding 83,310 for the set.
<b>Tasks</b>	Prediction (future disease codes)		Prediction task
<b>Comparison to</b>	PCA (principal component analyzed) / RawFeat (patient representation based on the original descriptors after removal of the frequent and rare variables) / K-Means / GMM (Gaussian mixture model) / ICA (independent component analysis).		<b>Supervised representation evaluation:</b> BoW / SDAE-BoW / doc2vec / [doc2vec, SDAE-BoW] / BoCUI . SDAE-BoCU <b>Feature analysis evaluation</b>
<b>Implementation available &amp; framework</b>	Python / Tensorflow <a href="https://github.com/natoromano/deep-patient/">https://github.com/natoromano/deep-patient/</a>		Python / Keras for the SDA <a href="https://github.com/MadhumitaSushil?tab=repositories">https://github.com/MadhumitaSushil?tab=repositories</a>

Tableau 2: Deep learning methods to learn patient representation using *transformers*

TAPER [12]	
<b>Date / Author / Jr</b>	August 2019 / Darabi & al. / Not submitted
<b>Purpose</b>	To build a unified vector representation using transformer networks and BERT language model to embed these data streams. This approach encodes a patient's visit data into a single distributed representation, which can be used for downstream tasks.
<b>Method:</b> architecture (auto-encoder / classifier)	<p>Training is 2 steps :</p> <p>(1) Skip-gram model using transformer networks to learn medical code representation</p> <p>(2) a BERT model is trained on medical notes and the resulting representations at a time step are summarized using autoencoder architectures</p> <p>The final representation for a patient is a concatenation of these two.</p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>(1)</p> </div> <div style="text-align: center;"> <p>(2)</p> </div> </div>
<b>Hyper-parameters</b>	<p>Training:</p> <p>Transformer encoder: dimension 128</p> <p>Self-attention module: 8 heads each 64 dimension.</p> <p>Network: adam optimizer / learning rate = 0.00025 / 2 layers / 50 epochs</p> <p>BERT (2 layers bidirectional GRU): d = 128 / lr : 0,5 every 50 epochs with initial lr = <math>10^{-3}</math> / teacher forcing ratio : 0,5 / Initialization with BioBert and finetune on MiMiC-III</p> <p>Classifier: 2 layers fully connected / fist layer ((dcode+denc+ddemographics)/2) neurons / ReLU activation. 2<sup>nd</sup> layer mapping the number of class / 30 epochs / learning rate schedule decay: 0,1 every 10 epochs (setting used for all downstream tasks)</p>
<b>Data set:</b> Data description Type of data Pretreatments	<p>MIMIC-III ICU (using codes &amp; text) [20]</p> <p>Pre-processing: use of the Clinical Classification Software for ICD9-CM5 to group the ICD9 diagnosis codes into 231 categories. + others minor pre-process</p> <div style="text-align: center;"> </div>
<b>Tasks</b>	Predictions tasks: readmission task, mortality task, length of stay task, code prediction task
<b>Comparison to</b>	Med2Vec [21] / ClinicalBERT [22] / Time-aware Embedding [23] / Patient2Vec [16] / DeepR [4]
<b>Implementation available &amp; framework</b>	Python / Pytorch <a href="https://github.com/sajaddarabi/TAPER">https://github.com/sajaddarabi/TAPER</a> (empty)

Tableau 3: Deep learning methods to learn patient representation using **transformers**

BEHRT [11]	
<b>Date / Author / Jr</b>	July 2019 / Li & al. / Not submitted
<b>Purpose</b>	A deep neural sequence transduction model for EHR capable of multitask prediction and disease trajectory mapping. With a flexible architecture that is capable of capturing more modalities of EHR data (4 key embeddings - diseases, age, segment and position).
<b>Method:</b> architecture (auto-encoder / classifier)	Transformer-based Model for EHR Feedforward neural networks to model time sensitive EHR data by examining the various forms of sequential order within the data. The model provides disease sequences, also give it with a precise <b>sense of timing</b> of events, and data about the delivery of care. Pre-training (same approach for BERT): masked language models (MLM) <b>Self-attention mechanism:</b> capturing non-temporal/non-directional relationships among diseases. To train a predictor, we feed the output from BEHRT to a single <b>feed-forward classifier layer</b> and train it three separate time for each of the three tasks.
<b>Hyper-parameters</b>	Optimal hyperparameters found with bayesian optimisation 6 layers, 12 attention heads, intermediate layer size of 512, and hidden size of 288
<b>Data set:</b> Data description Type of data Pretreatments	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>Clinical Practice Research Datalink (CPRD) contains longitudinal primary care data from a network of 674 general physician which is linked to secondary care and other health and administrative databases. Covers 35 million patients. CPRD is broadly representative of the population by age, sex, and ethnicity. It has been extensively validated and is considered as the most comprehensive longitudinal primary care database. HES contains data on hospitalizations, outpatient visits, accident and emergency for all admissions to NHS hospitals.</p> </div> <div style="width: 50%;"> <p><b>Pretreatment:</b> Started on 8 million patients. kept individuals who have at least 5 visits in their EHR</p> <p>Then forming sequences only with the diagnosis and age subset of the record.</p> </div> </div>
<b>Tasks</b>	Multitask prediction and disease trajectory mapping. (1) Prediction of concepts in the next visit (T1) (2) Prediction of the occurrence of diseases in the next 6 months (T2) (3) Prediction of the occurrence of diseases in the next 12 months (T3).
<b>Comparison to</b>	DeepR (was used bayesian optimisation to find the optimal hyperparameters) RETAIN (was used bayesian optimisation to find the optimal hyperparameters)
<b>Implementation available &amp; framework</b>	Not found

Tableau 4: Deep learning methods to learn patient representation using recurrent neural network (RNN) and self-attention

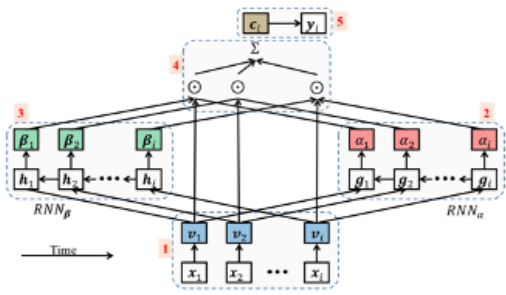
	RETAIN [9]	Patient2Vec [16]
<b>Date / Author / Jr</b>	February 2017 / Choi & al. / Accepted NIPS	2018 / Zhang & al. / IEEE
<b>Purpose</b>	Reverse Time Attention neural network model (RETAIN) is an end-to-end interpretable model, apply on EHR data, retrieving codes for prediction.	<i>Patient2Vec</i> , to learn an interpretable deep representation of longitudinal EHR data using hierarchical attention mechanism (RNN network) allowing the authors to directly interpret the weights of clinical events (extracted from codes)
<b>Method:</b> architecture (auto-encoder / classifier)	<p>Unfolded view of RETAIN's architecture: Given input sequence <math>x_1, \dots, x_i</math>, we predict the label <math>y_i</math>.</p> <p>Step 1: Embedding,</p> <p>Step 2: generating <math>\alpha</math> values using <math>RNN_{\alpha}</math></p> <p>Step 3: generating <math>\beta</math> values using <math>RNN_{\beta}</math></p> <p>Step 4: Generating the context vector using attention and representation vectors</p> <p>Step 5: Making prediction</p> <p>Note that in Steps 2 and 3 we use RNN in the reversed time.</p> 	<p>RNN - GRU network</p> <p>Framework in 5 parts:</p> <ol style="list-style-type: none"> <li>(1) Learning vector representations of medical codes</li> <li>(2) Learning within-subsequence self-attention</li> <li>(3) Learning subsequence-level self-attention</li> <li>(4) Constructing aggregated deep representation</li> <li>(5) Predicting outcome</li> </ol>
<b>Hyper-parameters</b>		No details in the article
<b>Data set:</b> Data description Type of data Pretreatments	<p>Electronic health records from Sutter Health</p> <p>Records, medication orders, procedure orders and problem lists, we extracted visit records consisting of diagnosis, medication and procedure codes.</p> <p>Use of existing medical groupers to aggregate the codes into input variables</p>	<p>EHR data from the University of Virginia Health System covering 75 months beginning in September 2010.</p> <p>2,343,651 inpatient and outpatient visits of 473,915 distinct patients, then extracted visit data with diagnosis, medication, and procedure codes.</p>
<b>Tasks</b>	Heart failure prediction	Prediction of future hospitalizations
<b>Comparison to</b>	LR, MLP, RNN (no comparison to other fellow models)	LR / MLP / bidirectional RNN with medical group embedding / forward RNN with medical vector embedding / bidirectional RNN with medical vector embedding / RETAIN
<b>Implementation available &amp; framework</b>	<p>Python / Theano</p> <p><a href="https://github.com/mp2893/retain">https://github.com/mp2893/retain</a></p>	<p>Python / Pytorch</p> <p><a href="https://github.com/BarnesLab/Patient2Vec">https://github.com/BarnesLab/Patient2Vec</a></p>



Tableau 5: Deep learning methods to learn patient representation using recurrent neural network (RNN)

	Doctor AI [8]	Health-ATM [13]
<b>Date / Author / Jr</b>	September 2016 / Choi & al. / Conference presentation 2016	September 2019 / Ma & al. / Conf SIAM
<b>Purpose</b>	Doctor AI is a RNN-based model that can learn efficient patient representation from a large amount of longitudinal patient records ( <b>codes</b> ) and predict future events of patients	Deep architecture that uncovers patients' comprehensive health conditions with irregular sampled to extracts multifaceted patient patterns with attentive and time-aware modulars (ATM)
<b>Method: architecture</b> (auto-encoder / classifier)	Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data (performs multilabel prediction on EHR codes). Learning an <b>effective vector representation</b> for the patient status at each timestamp. RNN with GRU	<p><b>RNN:</b> input = embedded clinical events Codes are laid out in a temporal order (each code represent a clinical event). Then applying the CBOV model of word2vec. bi-directional GRU</p> <p><b>CNN (time-aware ConvNet):</b> contains two layers; a multi-channel time-aware convolutional layer and a max-pooling layer.</p> <p><b>Prediction layer</b></p>
		<p>A embedding layer B hidden layer of GRU C attention weights D "context" vector E time-aware ConvNet F prediction</p>
<b>Hyper-parameters</b>	Softmax layer for code prediction (diagnosis and medication) / loss function: cross entropy ReLU to predict the time duration until next visit. / loss function: squared loss	Softmax function to normalize the time-aware weight Activation function: ReLU Multiple filters to generate n filter maps, and then apply a max-pooling layer subsequently to aggregate the representation. Window size of 15 for word2vec and train medical code vectors of 100 dimensions on each training data. The hidden layer size of GRU is 50 Number of filters for CNN is 64 and each filter window size is 51 Max-pooling layer following the convolution with pooling size (3,1). Dropout rate is 0.5. Training is done through Adam at learning rate 0.0002 with shuffled mini-batches (batch size 100).
<b>Data set:</b> Data description Type of data Pretreatments	Health Palo Alto Medical Foundation. Inputs: ICD-9 codes, medication codes, and procedure codes.	Two datasets - congestive heart failure (CHF) cohort extracted from a proprietary real world EHR data (SNOW dataset) - a publicly available artificial dataset (EMRbots). SNOW: records from 319,650 patients over 4 years
<b>Tasks</b>	Prediction (future physician diagnoses and medication orders)	Prediction (Risk prediction)
<b>Comparison to</b>	LR / MLP	Logistic Regression CNN (SDM&NIPS) RETAIN
<b>Implementation available &amp; framework</b>	<a href="https://github.com/mp2893/doctorai/">https://github.com/mp2893/doctorai/</a> Python / Theano	Theano (no implementation found)

Tableau 6: Deep learning methods to learn patient representation using **recurrent neural network (RNN)**

	DeepCare [5]	Darabi2 [17]
<b>Date / Author / Jr</b>	April 2017 / Trang & al. Accepted at JBI (Journal of Biomedical Informatics)	October 2019 / Darabi & al. / Not submitted
<b>Purpose</b>	DeepCare is an <i>end-to-end</i> deep dynamic memory neural network for personalized healthcare (Long Short-Term Memory (LSTM) architecture with attention mechanism)	Patient embedding using both clinical time-series signals and concepts using multi-modal data, namely, clinical concepts and <b>vital signals</b> This paper introduced an unsupervised patient status vector embedding scheme for EHR patient longitudinal data using a two-step autoencoding step.
<b>Method:</b> architecture (auto-encoder / classifier)	Deep dynamic memory neural network (LSTM) Input = information extracted from admissions Outputs = illness states at the time of admission Vector space to represent discrete elements (embedding) Pretraining can be done by unsupervised learning on unlabeled data	Was proposed a two-step unsupervised finetuning task for embedding patient data: (1) a single visit autoencoding step followed (Medical Concept Embedding & Clinical Time Series Embedding) ( <b>RNN-GRU</b> ) (2) forecasting autoencoding
<b>Hyper-parameters</b>	Disease progression and intervention recommendation task : M (embedding dimensions) = 30 and K (hidden units) = 40 / Prediction task : M = 10 and K = 20 / Mini-batch of 16 sequences / epochs before updating = 5 / epoch = 200 / Learning rate smaller than 0.0001	MSE loss Training the models: Adam / batch size = 64 Connected layer with ReLU as activation and dropout = 0.1
<b>Data set:</b> Data description Type of data Pretreatments	Data are from a large regional Australian hospital.	Collaborative Research Database v2.0 (eICU) 200,000 Intensive Care Unit (ICU) records collected from over 250 hospital sites in the United States, between 2014 and 2015 (multi-modal data, namely, clinical concepts + <b>vital signals</b> ).
<b>Tasks</b>	Prediction (predicts future medical outcomes)	
<b>Comparison to</b>	No comparison to other fellow models (only Classique SVM, Random Forest, Plain RNN, LSTM) BoW+LR (bag-of-words representation and regularized logistic regression)	
<b>Implementation available &amp; framework</b>	<a href="https://github.com/trangptm/DeepCare">https://github.com/trangptm/DeepCare</a>	Python / Pytorch 1.0 <a href="https://github.com/sajaddarabi">https://github.com/sajaddarabi</a> (No implementation found in author's GitHub)

Tableau 7: Deep learning methods to learn patient representation using *convolutional neuronal network (CNN)*

	DeepR [4]
<b>Date / Author / Jr</b>	January 2017 / Nguyen & al. / IEEE
<b>Purpose</b>	DeepR ( <u>D</u> eep net for medical <u>R</u> ecord) is an <i>end-to-end</i> deep learning system that learns to extract features from medical records (patient representation) and predicts future risk automatically
<b>Method:</b> architecture (auto-encoder / classifier)	<p>DeepR transforms a record into a sequence of discrete elements separated by coded time gaps and hospital transfers = “sentence” of multiple phrases separated by special “words” that represent time gap. On top of the sequence is a convolutional neural net that detects and combines predictive local clinical motifs to stratify the risk.</p> <p>(1) embedding layer à CNN (ReLU) / Max-pooling            (2) pooling à convolutional layer            (3) classifier à linear classifier / loistic regression or a non-linear parametric classifier /neural network</p> <p>Training: pre-training with <i>word2vec</i></p>
<b>Hyper-parameters</b>	<p>Found by grid-search:</p> <p>M (embedding dimension) = 100, d (kernel window size) = 1, motif size = 3, 4 and 5, n = 100 number of epochs = 10, mini-batch size = 64, <math>l_2</math> regularization <math>\lambda = 1.0</math>.</p>
<b>Data set:</b> Data description Type of data Pretreatments	<p>Data are from a large private hospital chain in Australia</p> <p>590546 records, each one admission defined by an admission time and a discharge time (300000 unique patients).</p>
<b>Tasks</b>	Prediction (predict unplanned readmission after discharge)
<b>Comparison to</b>	BoW+LR (bag-of-words representation and regularized logistic regression)
<b>Implementation available &amp; framework</b>	<p>No implementation found</p> <p>Python / TensorFlow</p>

## Annexe 4 : Table des figures et tableaux

<i>Figure 1 : Réseau neuronal avec une couche d'entrée, une couche de sortie et deux couches d'unités cachées</i>	<i>P. 4</i>
<i>Figure 2 : Représentation schématique de l'architecture d'un auto-encodeur (AE)</i>	<i>P.8</i>
<i>Figure 3 : Représentation schématique de l'architecture d'un auto-encodeur de débruitage (DAE)</i>	<i>P.9</i>
<i>Figure 4 : Représentation schématique de l'architecture d'un auto-encodeur de débruitage superposé (SDAE)</i>	<i>P.10</i>
<i>Figure 5 : Échantillon réduit</i>	<i>P.11</i>
<i>Figure 6 : Graphique en violon de la distribution des âges</i>	<i>P.12</i>
<i>Figure 7 : Graphique en violon de la distribution de l'âge par sexe pour la population pédiatrique</i>	<i>P.13</i>
<i>Figure 8 : Représentation graphique des données</i>	<i>P.14</i>
<i>Tableau 1 : Résumé du prétraitement des données et de type d'entrée des modèles de l'état de l'art</i>	<i>P.6</i>
<i>Tableau 2 : Résumé du prétraitement des données et de type d'entrée des modèles de l'état de l'art</i>	<i>P.6</i>
<i>Tableau 3 : Résumé du prétraitement des données et de type d'entrée des modèles de l'état de l'art</i>	<i>P.6</i>
<i>Tableau 4 &amp; 5 : Démographie et âge de la population</i>	<i>P.12</i>
<i>Tableau 6 &amp; 7 : Démographie et âge de la population pédiatrique</i>	<i>P.13</i>
<i>Tableau 8 : répartition des codes UMLS CUI</i>	<i>P.13</i>
<i>Tableau 9 : Réhospitalisations</i>	<i>P.14</i>
<i>Tableau 10 : Résultats des modèles de références pour la base de donnée totale</i>	<i>P.15</i>
<i>Tableau 11 : Résultat du SDEA pour la base de donnée totale</i>	<i>P.15</i>
<i>Tableau 12 : Résultat du SDEA pour les données pédiatriques</i>	<i>P.15</i>

# SERMENT D'HIPPOCRATE

**Au moment d'être admis(e) à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.**

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans **aucune discrimination selon leur état ou leurs convictions**. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas **usage de mes connaissances contre les lois de l'humanité**.

**J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.**

Je ne tromperai **jamais leur confiance** et **n'exploiterai pas le pouvoir hérité** des circonstances pour forcer les consciences.

**Je donnerai mes soins à l'indigent et à quiconque me les demandera**. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

**Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés**. **Reçu(e) à l'intérieur des maisons, je respecterai les secrets des foyers** et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

**Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission**. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

**J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.**

**Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré(e) et méprisé(e) si j'y manque.**



# Résumé

## **Objectif :**

Construire une modélisation du dossier patient par deep learning réutilisable pour différentes sous-tâches à partir de l'entrepôt de données cliniques de l'hôpital Necker puis l'appliquer à une sous-tâche de test : la réhospitalisation des patients en moins de 30 jours.

## **Matériaux et méthodes :**

Nous avons premièrement conçu un alignement (mapping) des codes CIM-10, des codes LOINC de biologie avec les CUI de l'UMLS. Dans un second temps nous avons utilisé les concepts précédemment extraits des notes médicales puis nous les avons alignés avec les CUI de l'UMLS. La modélisation a été construite avec une structure de deep learning semi-supervisée : les stacked denoising auto-encoder (SDAE). Nous avons testé le SDAE sur différents échantillons de la base de données (1000, 10000, toutes les données ainsi qu'un sous-échantillonnage).

## **Résultats :**

Le SDAE a obtenu de meilleurs résultats que l'ensemble des échantillons de références, à l'exception du sous-échantillonnage où le DEA a obtenu de meilleurs résultats.

## **Discussion :**

Nous montrons dans cette étude que la modélisation du dossier patient d'un hôpital pédiatrique par deep learning permet après application à une sous tâche de classification de prédire les patients réhospitalisé en moins de 30 jours. Nous montrons aussi que la précision de prédiction augmente sensiblement lorsque l'algorithme de deep learning (SDAE) est appliqué uniquement à la base pédiatrique, plus petite et plus spécifique. Plusieurs améliorations semblent envisageables et d'autres sous-tâches doivent être testées.

## **Mots clés**

Medical Informatics; Electronic health records; Deep learning; Unsupervised learning; Patient representations