AAP Contrats doctoraux en Intelligence artificielle

Cofinancé par l'ANR

DYSNATREM.IA : Artificial intelligence to prevent dysnatremias

1. DESCRIPTION OF THE PHD THESIS PROJECT

1.1 OBJECTIVES OF THE PROJECT BASED ON THE CURRENT STATE OF THE ART

In the medical field, application of artificial intelligence to the analysis of hydroelectrolytic disorders is still in its infancy and has received very few contributions [1]. However, such an analysis is clinically relevant because of potential exhaustivity of hydroelectrolytic disorders diagnosis and medical time saving. Currently, since the famous book of Homer Smith, the fish and philosopher, control of the "milieu interieur" is known to be essential to all cellular and organ functions. More generally, homeostasis is a key concept of physiology, pathology and medical decision making. As an example, Hyponatremia, a common hydroelectrolytic disorder, affects more than 10% of hospitalized patients. To analyze such disorders, the current medical approach is analytic and physician supported. Unfortunately, this leads to a scotomization of the anomalies, especially because of the complexity of the physiopathological mechanisms underlying the disorders. Moreover, kidney physiology regarding water disorders can be perceived counterintuitive by students and MD. But Kidney is the key regulator of the hydroelectrolytic equilibrium and AI technologies are now able to predict up to 48h in advance occurrences of acute kidney injury [2]. We therefore think that AI can play a significant role to help non-specialist physicians to analyze and predict the hydroelectrolytic disorder of their patients. More precisely, in our opinion, an AI conjoint automatic and systematic analysis of the blood ionogram and of biomarkers of the renal function (creatininémia and uremia), which are among the most frequently prescribed biological tests,should be able to improve patients' medical support through a better diagnosis rate, data-driven care, reduced morbidity, reduced length of stay and finally reduced costs [3].

The goal of this thesis is thus to create a medical decision making tool designed to facilitate dysnatremia (abnormal natremia) detection, i.e. hyponatremia and hypernatremia, that is the most frequent hydroelectrolytic disorder. Our goal is to prevent the development of the severe forms of these disorders to reduce the mortality associated with them. The analysis will use the results of biological tests integrated with the analysis of the clinical history of the patients and the drugs prescribed to the patient. For the purpose of this thesis, we will rely on probabilistic graphical models (PGM) like Bayesian networks [5] or causal networks [6] to exploit them. We chose to use these PGMs because, unlike black-box models such as deep convolutional networks [7], they can easily be interpreted and validated by medical experts. This PhD thesis is therefore an AI exploratory research project focusing on health diagnostic assistance problem solving.

1. *PubMed research 08/13/2019 « Hypernatremia OR Hyponatremia AND Artificial Intelligence »: Tate KE, Gardner RM, Weaver LK. A computerized laboratory alerting system. MD Comput. 7(5):296-301. 1990.*

2. *Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, Connell A, Hughes CO, Karthikesalingam A, Cornebise J, Montgomery H, Rees G, Laing C, Baker CR, Peterson K, Reeves R, Hassabis D, King D, Suleyman M, Back T, Nielson C, Ledsam JR, Mohamed S. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature, 572(7767):116-119. 2019.*
3. *MacLeod, A. NCEPOD report on acute kidney injury—must do better. Lancet 374, 1405-1406. 2009.*
4. *Dasta JF, Kane-Gill S. Review of the Literature on the Costs Associated With Acute Kidney Injury. J Pharm Pract. 32(3):292-302. 2019.*
5. *Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman Publishers, 1988.*
6. *Pearl J. Causality. Vol. 2. Cambridge University Press, 2009.*
7. *Goodfellow I, Bengio Y, Courville A. Deep Learning. The MIT Press, Cambridge. 2017.*

1.2 METHODOLOGY

As mentioned above, for the purpose of this project, we will rely on PGMs because they can easily be interpreted and validated by medical experts. This is a feature of utmost importance in the medical field (and more generally in critical/high stakes applications). This explains why PGMs have been used in different medical areas for many years [8,9]. Bayesian and causal networks can encode conditional dependencies/causal relations among various physiological, pathological and etiological states of a patient. Particularly, several physical laws are known to drive homeostasis including hydroelectrolytic status and deterministic causal links and organ cross-talking in the complex system that is Human are essential to be captured. Beside, their predictions can be interpreted: it is actually possible to determine which variables led these models to predict these states and their possible anomalies [10,11]. It should also be noted that causality, which plays a central role in this thesis, cannot be currently captured by any other mathematical tool than causal networks [12] and is currently a very active field of research in the AI community [13] and well in the network physiology community [14].

The first step in order to exploit such PGMs is to construct them from both data and expert knowledge. Learning from data is a central topic in AI and many algorithms do exist for this purpose [15]. However, in the context of this project, this raises a first issue: that of determining the learning dataset. The PhD student will therefore have to create one from the aggregation of clinical, biological and pharmaceutical data available on a given period (of at least one year).

In a second step, PGM structure and parameter learning will be performed on those data. Even though there exists a large literature on this topic, several issues will need to be fixed. First, mixed probability distributions involving both discrete and continuous variables are very challenging to learn and are still actively investigated in the AI community [16]. Second, deterministic or quasi-deterministic relationships between random variables due to physical relations (e.g., anions and cations, natremia and chloremia, etc.) rule out the faithfulness property on which the majority of learning algorithms rely [17], which prevents them from learning valid PGMs. Here again, this issue needs to be dealt with. Third, non-stationary relations need also be taken into account. Finally, learning causal networks while taking into account all these constraints requires the development of new sophisticated algorithms.

Once PGMs are available, we plan to exploit them in a third step to predict dysnatremia. This also raises technological challenges, notably that of performing fast inferences in a temporal (dynamic) setting [18], especially for causal queries, and of the interpretations of their results [11]. But this also raises the issue of how corrective measures can be taken, which leads to the problem of causal inference (do-calculus [6]).

Finally, the efficiency and effectiveness of all these models and algorithms will be assessed through a clinical trial. The best models will be integrated in a medical decision making prototype designed for the management and prevention of dysnatremia. Comparisons with the traditional medical approach for the prevention of hydroelectrolytic disorders will be performed both on a medical and medico-economic points of view. This will enable us to prove the usefulness of exploiting AI at the bedside to improve the care of patients and reduce the iatrogeny.

8. *Beinlich IA, Suermondt HJ, Chavez RM, Cooper GF. The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks. Proceedings of the European Conference on Artificial Intelligence in Medicine, pp. 247-256. Springer-Verlag, 1989.*
9. *Spiegelhalter DJ, Cowell RG. Learning in probabilistic expert systems. In Bayesian Statistics (JM Bernardo, J0 Berger, AP Dawid and AFM Smith, eds.) 4:447-466. Clarendon Press, Oxford. 1992*
10. *Liu Q, Ihler A. Variational algorithms for marginal MAP. Journal of Machine Learning Research, 14:3165-3200. 2013.*
11. *Yuan C, Lim H, Lu TC. Most relevant explanation in Bayesian network. Journal of Artificial Intelligence Research, 42(1):309-352. 2011.*
12. *Pearl J, Mackenzie D. The book of why: the new science of cause and effect. Basic books, 2018.*
13. *Pearl J. Causal Analysis in The Theory and Practice. http://causality.cs.ucla.edu/blog/*
14. *Bashan A, Bartsch RP, Kantelhardt JW, Havlin S, Ivanov PCh. Network physiology reveals relations between network topology and physiological function. Nat Commun. 3:702. 2012.*
15. *Koller D, Friedman N. Probabilistic Graphical Models: Principles and Techniques. The MIT Press. 2009.*
16. *Salmerón A, Rumí R, Langseth H, Nielsen T.D., Madsen AL. A Review of Inference Algorithms for Hybrid Bayesian Networks. Journal of Artificial Intelligence Research, 62(1):799-828. 2018.*
17. *Mabrouk A, Gonzales C, Jabet-Chevalier K, Chojnaki E. An Efficient Bayesian Network Structure Learning Algorithm in the Presence of Deterministic Relations. Proceedings of the European Conference on Artificial Intelligence, pp. 567-572. 2014.*
18. *Chopin M, Wuillemin PH. Optimizing the triangulation of Dynamic Bayesian Networks. Proceedings of Probabilistic Graphical Models, pp. 73-80, 2010.*

1.3 WORK PLAN

The methodology described above suggests splitting the project into five tasks. Below we provide some additional details about these tasks.

Task 0 [T0-T6]: State of the art review on both probabilistic graphical models and on network physiology focusing on renal function and hydroelectrolytic disorders (targeting dysnatremia prediction).

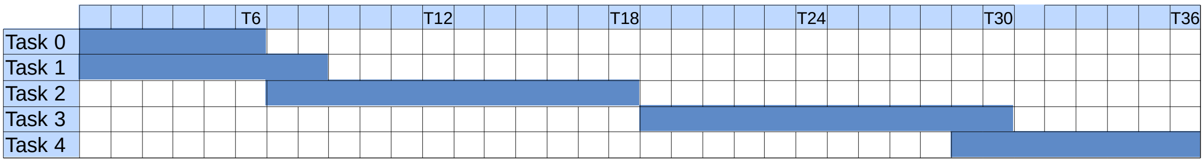
Task 1 [T0-T8]: Creation of the dataset. Queries will need to be sent to the data warehouse of the medical IT service of the Assistance Publique – Hôpitaux de Marseille (AP-HM). This will require regulatary process including the authorization to exploit health data from the AP-HM and the CNIL Health Data Access Commission, as well as setting up an anonymization procedure.

Task 2 [T6-T18]: PGM Learning from heterogeneous data (discrete/continuous). In order to perform supervised or semi-supervised learning, annotations (of states or pathologies to be identified and to be predicted) will rely on the encoding of the information systems medicalization program (PMSI) but also on the use of natural language processing on the patients discharge reports. Together, those will define the ground truth.

Task 3 [T18-T30]: Algorithms for predicting natremia anomalies, explaining their causes and determining the appropriate countermeasures. Here, it is important to note the causal dimension of these problems: they require solutions that are totally different from the usual probabilistic/statistical algorithms or even from deep learning models (see [6,12] for a thorough discussion on the reasons why the latter can provide strongly incorrect results). The algorithms developed in the project will be adapted to take into account both the prior knowledge on the physiology and physiopathology of the "milieu interieur" homeostasis, and to the knowledge on the technical characteristics of the automatons measuring the target parameters.

Task 4 [T28-T36]: clinical trial.

We plan to organize these tasks according to the following Gantt chart:



1.4 SUPERVISOR AND RESEARCH GROUP DESCRIPTION

This research project will be supervised by Stéphane Burtey, Stéphane Delliaux and Christophe Gonzales. The medical nephrological aspects will be carried out by Stéphane Burtey, the medical physiological aspect and medical-AI interface by Stéphane Delliaux while the AI aspect will be devoted to Christophe Gonzales.

The project is neither supported by other research funds nor is part of a larger research programme.

2. RECENT PUBLICATIONS

Publications related to the PGM dimension of the project:

- *Cortijo S, Christophe Gonzales. On conditional truncated densities Bayesian networks. International Journal of Approximate Reasoning. 92:155-174. 2018.*
- *Mabrouk A, Gonzales C, Jabet-Chevalier K, Chojnacki E: An Efficient Bayesian Network Structure Learning Algorithm in the Presence of Deterministic Relations. Proceedings of the European Conference on Artificial Intelligence, pp. 567-572. 2014.*
- *Hourbracq M, Wuillemin PH, Gonzales C, Baumard P. Learning and Selection of Dynamic Bayesian Networks for Non-Stationary Processes in Real Time. Proceedings of FLAIRS Conference, pp. 742-747. 2017.*
- *Gonzales C, Dubuisson S, Manfredotti CE. A New Algorithm for Learning Non-Stationary Dynamic Bayesian Networks With Application to Event Detection. Proceedings of FLAIRS Conference pp. 564-569. 2015.*

Publications related to the integrative physiological dimension of the project:

- *Delliaux S, Delaforge A, Deharo JC, Chaumet G. Mental Workload Alters Heart Rate Variability, Lowering Nonlinear Dynamics. Front Physiol. 2019 May 14;10:565.*
- *Pons JF, Haddi Z, Deharo JC, Charaï A, Bouchakour R, Ouladsine M, Delliaux S. Heart rhythm characterization through induced physiological variables. Sci Rep. 2017 Jul 11;7(1):5059.*
- *Delliaux S, Ichinose M, Watanabe K, Fujii N, Nishiyasu T. Cardiovascular responses to forearm muscle metaboreflex activation during hypercapnia in humans. Am J Physiol Regul Integr Comp Physiol. 2015 Jul 1;309(1):R43-50.*

Publications related to the nephrological aspects of the project:

- *Lano G, Sallée M, Pelletier M, Bataille S, Fraisse M, Berda-Haddad Y, Brunet P, Burtey S. Mean Platelet Volume Predicts Vascular Access Events in Hemodialysis Patients. J Clin Med. 2019 May 4;8(5). pii: E608. doi: 10.3390/jcm8050608 SMASH. PubMed PubMed ID 31060235 SMASH; PubMed Central PMCID: PMC6571831.*
- *Dou L, Poitevin S, Sallée M, Addi T, Gondouin B, McKay N, Denison MS, Jourde-Chiche N, Duval-Sabatier A, Cerini C, Brunet P, Dignat-George F, Burtey S. Aryl hydrocarbon receptor is activated in patients and mice with chronic kidney disease. Kidney Int. 2018 Apr;93(4):986-999. doi: 10.1016/j.kint.2017.11.010 SMASH. Epub 2018 Feb 1. PubMed PubMed ID 29395338 SMASH.*
- *Santana Machado T, Poitevin S, Paul P, McKay N, Jourde-Chiche N, Legris T, Mouly-Bandini A, Dignat-George F, Brunet P, Masereeuw R, Burtey S, Cerini C. Indoxyl Sulfate Upregulates Liver P-Glycoprotein Expression and Activity through Aryl Hydrocarbon Receptor Signaling. J Am Soc Nephrol. 2018 Mar;29(3):906-918. doi: 10.1681/ASN.2017030361 SMASH. Epub 2017 Dec 8. PubMed PubMed ID 29222397 SMASH; PubMed Central PMCID: PMC5827590.*
- *Burtey S, Sternberg D, Nguyen K, Philip N, Berland Y, Dussol B. Hypokalaemia and dysmorphia, is there a link? NDT Plus. 2009 Jun;2(3):222-4. doi: 10.1093/ndtplus/sfp009 SMASH. Epub 2009 Feb 4. PubMed PubMed ID 25983995 SMASH; PubMed Central PMCID: PMC4421201.*
- *Burtey S, Dussol B, Philip N, Berland Y. The Case: Hypocalcemia, chronic renal failure and dysmorphism. Kidney Int. 2008 Dec;74(11):1495-6. doi: 10.1038/ki.2008.430 SMASH. PubMed PubMed ID 19008915 SMASH.*

3. EXPECTED PROFILE OF THE CANDIDATE

We are looking for a highly motivated candidate with a strong background in mathematics and computer science. He/She should have recently completed a master's degree or should be about to complete it. The candidate should be interested in life science and medical practices since this is the scope of the project and he/she will often interact with physicians. The candidate should also demonstrate good programming skills, notably in C++ and Python, because the project will include developments using these languages. An experience or some understanding of probabilistic graphical models, e.g., of Bayesian networks, would be appreciated.

According to the interdisciplinary focus of the project and of the team the candidate is also expected to be open-minded, curious, motivated to learn a little bit about Human functioning and medical knowledge, and to be able to deal out of its condor space.

4. SUPERVISORS' PROFILE

Christophe Gonzales is full professor in computer science at the laboratoire d'informatique et systèmes (LIS), Aix-Marseille University. He has completed his PhD in decision theory at Sorbonne University in 1996 and, since, has conducted research activities in artificial intelligence, more precisely on graphical models for decision making, with a special focus on probabilistic graphical models. From 2005 to 2014, he has been head of the artificial intelligence MsC department at Sorbonne University. He is currently supervising one PhD thesis, which started on January, 1st 2020, and has supervised 8 other PhD theses.

Stephane Burtey is full professor of nephrology in Aix-Marseille University. He completed medical formation in Aix-marseille university in 2000 and PhD in biology of eukaryotes in 2006. He is in charge of the daycare and consultation unit in Nephrology center in hopital de la conception. He is also group leader of the NEMOT team, a joint INSERM-AMU team in the C2VN. He is a specialist of the milieu interieur focusing on the uremic toxins and their effects on the cardiovascular system. He has also been interested in hydroelectrolytic disorders for years. He currently supervises two PhD theses (November 2018 and November 2017) and had supervised 6 PhD theses. He currently supervises a MD thesis on hypernatremia and also supervised more than 10 MD thesis.

Stéphane Delliaux is assistant professor - full lecturer in physiology at Aix-Marseille University. He is an intensivist medical doctor (MD, 2003) and physiologist (PhD, 2006). He is in charge of the cardiorespiratory and exercise physiological tests at the North Hospital of Marseille (AP-HM). He is also researcher in the Dysoxia Team of the joint INSERM-AMU team that is the C2VN (Center for CardioVascular and Nutrition research, Aix-Marseille University). He focuses on physiological and medical states modeling and co-leads the creation of a federal structure merging numerical sciences and artificial intelligence forces to target health and medical challenges. He is currently preparing his leading research habilitation.

VISA DU RESPONSABLE DE L'INSTITUT ET DU DIRECTEUR DE LABORATOIRE CONCERNÉS

Visa du responsable de l'institut, NOM Prénom	Visa du directeur du laboratoire, NOM Prénom
Fait à Marseille, le	Fait à Marseille, le 20/05/2020
Signature	Signature Frédéric BECHET Directeur du LIS UMR 7020 AMU-CNRS