

Analyse exploratoire de données d'insulinothérapie du diabète de type 2

Etude de la variabilité des besoins insuliniques

A. Nourizadeh^{1,2,3}, F. Blanchard³, A. Ait-Younes³, B. Delemer², M. Herbin³

¹ Groupe Axon³, Projet Domocare,

² Service d'Endocrinologie-Diabète-Nutrition, CHU de Reims,

³ CReSTIC (EA 3804) Université de Reims Champagne-Ardenne

Abstract— The ultimate goal of diabetes treatment is to improve the health of diabetics as well as reduce the risk of macrovascular and microvascular associated with diabetes by maintaining the quality of life of the patients. For this purpose, we conducted a study on the needs assessment of insulin in the elderly type 2 diabetic patients to modelize insulin treatment from physiological parameters. We use statistical and learning methods to replace the empirical and typically pattern at the initiation of insulin therapy. This to assist in the adjustment of insulin treatment of type 2 diabetes. This study is part of a PhD thesis funded by the company AXON³ and conducted in Reims Hospital University and research center CReSTIC.

Résumé— Le but ultime du traitement du diabète est d'améliorer l'état de santé des diabétiques et de réduire les risques microvasculaires et macrovasculaires associés au diabète en préservant la qualité de vie du patient. Dans cet objectif, nous avons réalisé une étude sur l'évaluation des besoins insuliniques des diabétiques âgés de type 2 afin de modéliser le traitement par l'insulinothérapie à partir des paramètres physiologiques. Nous utilisons les méthodes statistiques et d'apprentissage afin de remplacer la méthode empirique et classique lors du début de l'insulinothérapie. Il s'agit d'aider à l'ajustement des traitements insuliniques du diabète de type 2. Cette étude est effectuée dans le cadre d'une thèse financée par la société AXON³ et menée au sein de CHU de Reims et du centre de recherche CReSTIC.

Keywords—diabète, sujet âgé, insulinothérapie, télémédecine, analyse multidimensionnelle, clustering, arbre de décision

I. INTRODUCTION

Le diabète est une maladie chronique caractérisée par une « hyperglycémie résultant d'un défaut de sécrétion de l'insuline ou de son action, ou des deux conjuguées ». Il existe deux types de diabète : le type 1 et le type 2. Cette étude concerne le diabète de type 2 qui se manifeste en général après 40 ans avec deux phénomènes parfois concomitants : un manque d'insuline et une insulino-résistance.

En septembre 2012, on dénombre 347 millions de diabétiques dans le monde. En 2030, l'OMS estime qu'il y aura 438 millions de personnes touchées par cette maladie. En France, plus de 3,5 millions de personnes sont diabétiques dont 26% ont plus de 75 ans [1]. La région Champagne-Ardenne est particulièrement atteinte par cette épidémie [2]. La prévalence du diabète dans cette région est de 8,7% par an contre 5% en

France. L'étude présentée dans cette communication a débuté en Champagne-Ardenne en 2012.

Avec l'âge et le manque d'autonomie, les risques de complication augmentent. Les accidents cardiovasculaires et cérébraux constituent la principale cause de décès des diabétiques de type 2. Un suivi constant est alors impératif. Le passage à l'insulinothérapie change la vie du patient et de son entourage, l'hypoglycémie constitue alors un risque majeur qu'il convient de prévenir. Un accompagnement du patient âgé sous insuline est indispensable. Dans le contexte du maintien à domicile de la personne âgée diabétique, une éducation et une aide à l'insulinothérapie sont nécessaires aussi bien pour le patient et son entourage que pour l'équipe de soin et les professionnels de santé qui doivent pouvoir suivre l'évolution des pathologies des patients et adapter les traitements.

Des outils d'aide à l'insulinothérapie sont proposés ou en développement avec plusieurs objectifs :

- limiter le nombre d'interventions peu utiles pour le personnel médical et paramédical spécialisé,
- éviter des déplacements superflus à des patients fragiles,
- renforcer l'efficacité des interventions médicales pour mieux suivre et ajuster les comportements face à la maladie ou face à une thérapeutique,
- conserver un lien avec les patients en facilitant des interventions urgentes sans couper le patient du réseau de soins local.

Ces nouveaux outils sont indispensables au soutien à domicile des personnes fragiles ou isolées.

Les systèmes d'aide à l'insulinothérapie destinés aux personnes âgées doivent leur permettre d'augmenter leur autonomie vis-à-vis de la maladie et leur qualité de la vie au sein de leur propre domicile. Généralement, ces systèmes sont constitués de cinq composantes (Figure 1) avec :

- 1) la mesure et la collecte de données physiologiques telles que la glycémie et des informations comportementales telles que l'apport alimentaire et l'activité physique.
- 2) un dossier médical électronique pour la saisie des données et la transmission à distance par internet, GPRS, 3G...

- 3) un ensemble de protocoles pour l'analyse des données à distance.
- 4) une variété d'outils de communication pour permettre un dialogue efficace entre les patients et l'équipe de soins, tel que la visiophonie, l'email...
- 5) un système de retour automatique ou semi-automatique de signalisation, d'aide à la décision et d'assistance à la téléconsultation : tel que l'email, l'appel téléphonique... [3]

Dans la littérature, la plupart des systèmes proposés concernent le diabète de type 1 [4-11]. Hélas, les algorithmes d'insulinothérapie sont alors inadaptés au diabète type 2. La complexité de manipulation des différents dispositifs comme le lecteur de glycémie avec la transmission des données, par exemple par smartphone, les rendent peu utilisables pour des personnes âgées. Ces dispositifs nécessitent le développement d'une interface simplifiée, spécifique et très intuitive.

Les systèmes qui s'adressent aux diabétiques de type 2 sont principalement utilisés avec une seule injection par jour et les patients sous multi-injections ne peuvent pas en bénéficier. Ces systèmes sont encore des prototypes, ils n'ont pas été validés sur le long terme avec des effectifs significatifs, leur coût total est rarement évalué, ils manquent d'interfaces simples, adaptées et faciles d'accès pour des personnes âgées.

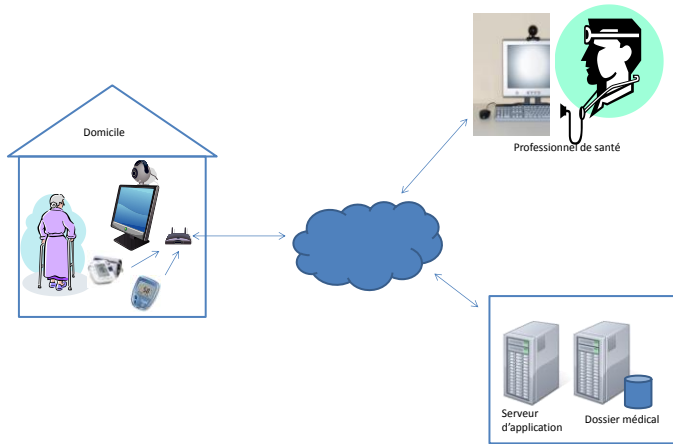


Fig. 1. Schéma général d'un système télésanté appliqué au diabète

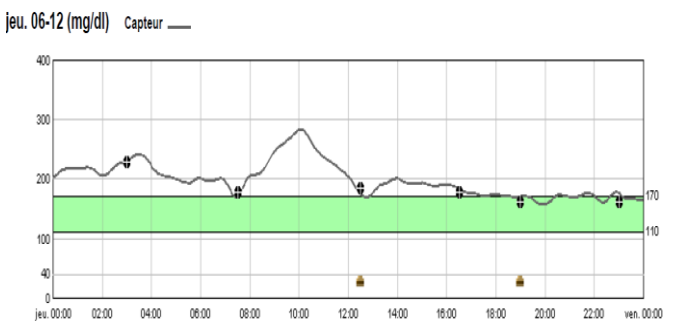


Fig. 2. Enregistrement continu du glucose par le capteur glycémique sur une période de 24 heures

Dans cette communication nous allons contribuer à l'émergence d'un système d'aide à l'insulinothérapie des diabétiques âgés de type 2. Les données recueillies lors de l'étude décrite dans cette communication sont complexes. Leurs natures multivariées et hétérogènes ainsi que l'absence d'hypothèse *a priori*, nous ont conduit à utiliser des techniques empruntées au *data mining* pour explorer ces données. Pour cela nous proposons d'étudier d'abord les spécificités des besoins insuliniques des patients diabétiques âgés de type 2. Dans la partie II, nous décrirons les données collectées lors de cette étude. Dans la partie III, nous présentons les différentes méthodes de classification et les potentielles typologies qui en découlent. La partie IV est consacrée à la conclusion et à la discussion des résultats.

II. DESCRIPTION DE L'ETUDE

Cette étude prospective et mono-centrique a pour objectif de mieux comprendre l'insulinothérapie du diabète de type 2 chez le sujet âgé sous schéma de multi-injection insulinaire (4 ou 5 injections par jour). L'étude est préalable à la construction d'un système d'aide à l'insulinothérapie.

Nous étudions les traitements insuliniques prescrits à 53 patients (M/F=24/29), âgés (65-88 ans), diabétiques de type 2 et hospitalisés sur périodes de 5 jours pour mise en place ou adaptation de leur traitement insulinaire en raison d'un déséquilibre persistant du diabète. Cette étude a été réalisée entre janvier 2012 et mars 2013 à l'hôpital de Reims (CHU Debré, service de diabétologie).

Pour construire un modèle d'insulinothérapie individualisé pour chaque patient, il faut tenir compte de l'alimentation (régime alimentaire en fonction de répartition glucidique de chaque repas selon les goûts et les habitudes de chaque personne qui est défini par la diététicienne), de l'exercice physique, des variabilités glycémiques aiguës, de l'objectif glycémique fixé par le diabétologue et de la protection vis-à-vis des épisodes graves (hypoglycémie ou hyperglycémie) et des facteurs modifiant les besoins insuliniques d'un patient à l'autre. Dans 10 cas, la prescription insulinaire a été validée sur l'ensemble de la durée de la journée et de la nuit par un enregistrement continu du glucose réalisé à l'aide d'un capteur glycémique (voir la figure 2)

Pour analyser les besoins en insuline, nous avons recueilli les données pour chaque patient. Dans cette étude nous utilisons les variables suivantes :

- *age* : Âge
- *poids* : Poids en kg
- *duree* : Durée d'évolution du diabète en années
- *HBA1C* : Moyenne des glycémies de trois derniers mois en %
- *MDRD* : Facteur de l'insuffisance rénale en ml/min
- *IDMAVC* : Présence ou absence de macro-angiopathie
- *glu* : Apport glucidiques de chaque repas en gr
- *basal* : Doses d'insuline lente en unités injectées

- *bolus* : Doses d'insuline rapide en unités injectées
- *imc* : IMC : Indice de Masse Corporelle est calculé : poids/ (taille)² soit poids en kg et taille en m
- *ukg* : Doses totales journalières en unités injectées par kg
- *doseTotal* : nombre total d'unités d'insuline injectées

III. CONSTRUCTION DE TYPOLOGIES DE PRESCRIPTION INSULINIQUE

Les données recueillies reflètent la complexité de la prescription insulinique dans le cas du diabète du type 2. La méthodologie d'analyse de ces données relève des techniques de *data mining* et se décompose en plusieurs étapes. La première est une phase de classification automatique non supervisée, ou *clustering*, effectuée sur le jeu de données afin de partitionner l'ensemble des patients (individus statistiques) en fonction des prescriptions faites à l'issue de leur hospitalisation. L'idée est donc de déterminer des classes de cas, dépendantes des deux variables de prescription d'insuline lente (*basal*) et d'insuline rapide (*bolus*), et de paramètres qui définissent l'espace paramétrique dans lequel nous appliquons les algorithmes de partitionnement.

La seconde phase est une étape de description des classes obtenues. Cette tâche est nécessaire pour comprendre les différences entre les groupes d'individus. La discrimination de ces groupes, à partir des autres variables, présente un double objectif : permettre d'interpréter la partition et donner du sens à ces regroupements, et cibler les variables externes les plus discriminantes.

Enfin la troisième phase repose sur la conception d'un mécanisme d'affectation des individus dans ces classes de prescription sans connaître, naturellement, le résultat des prescriptions insuliniques fournies à l'issue de l'hospitalisation. L'idée est donc de construire un algorithme de classement des individus utilisant les variables non issues des prescriptions des médecins.

A. Clustering

On considère que l'ensemble des individus de l'échantillon sont décrits à l'aide de plusieurs variables. La finalité des algorithmes de *clustering* est de déterminer une partition de cet ensemble en fonction de critères d'optimisation qui leurs sont propres (voir [14, 15], pour plus de détails sur les algorithmes de *clustering* de référence).

Cette partition regroupe les individus par similarité. Il convient donc de définir comment sont décrits les individus et comment sont mesurées les dissimilarités. L'espace paramétrique de description des individus, pour cette étape de *clustering*, est défini par :

- les variables de prescription insuliniques *basal* et *bolus*, décrivant la prescription faites par les médecins,
- des variables exogènes décrivant l'individu lui-même : *age*, *poids*, *imc*, *duree*.

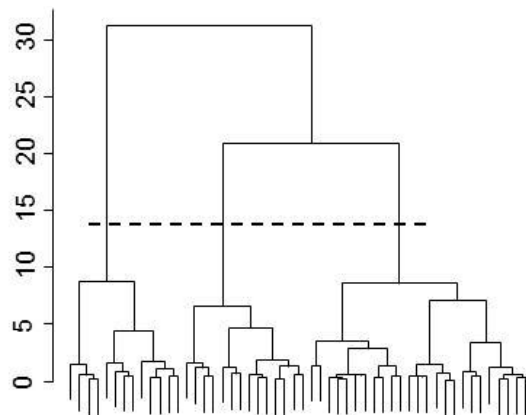


Fig. 3. Dendrogramme obtenu à l'issue de la classification ascendante hiérarchique.

On obtient ainsi une description des patients tenant compte de la prescription médicale à l'issue de l'hospitalisation mais aussi des paramètres intrinsèques aux patients (physiologiques). La mesure de dissimilarité que nous avons retenue, en première intention, est la distance euclidienne. Nous ne l'utilisons toutefois pas directement dans l'espace paramétrique, mais effectuons préalablement une analyse en composantes principales afin de projeter les individus dans un espace de dimension réduite. L'ACP présente aussi l'intérêt de fournir des composantes dé-corrélées. L'examen de l'ébouil des valeurs propres montre qu'il est possible de ne retenir que deux composantes (technique du « coude »). La dissimilarité entre individus est enfin calculée par distance euclidienne entre les projections de ces individus sur les deux premiers axes principaux.

Après avoir défini l'espace paramétrique et la mesure de dissimilarité entre individus, une première passe de classification ascendante hiérarchique (CAH) (méthode de Ward) nous permet de fixer le nombre de classes à retenir. La lecture du dendrogramme de la figure 3 nous suggère assez nettement que le nombre de classes à retenir est de 3.

Plusieurs algorithmes de clustering ont été comparés : l'algorithme des *k-means*, la *classification ascendante hiérarchique* et l'algorithme des *k-medoids*. Les trois algorithmes classiques donnent des résultats concordants (81,1% de classements concordants). La partition obtenue que nous proposons est donc constituée de trois classes d'effectifs respectifs 17, 20 et 16. L'étape suivante consiste alors à décrire ces classes et à en extraire les principales caractéristiques.

B. Description des classes

Les algorithmes de *clustering* utilisés fournissent une partition en classes. Le tableau I présente quelques statistiques descriptives élémentaires de ces classes. Le tableau I fournit un « individu moyen » pour chaque classe qui est présenté dans la colonne *Moyenne*. L'algorithme des *k-medoids* se distingue de celui des *k-means* en utilisant des individus comme centres de classes. On obtient donc des individus représentatifs des classes issus de l'échantillon étudié et non des éléments fictifs ou « virtuels ». Ces individus représentatifs sont nommés *Medoid* dans le tableau I.

Tableau I. Statistiques descriptives des trois classes.

Classe 1

Variable	Moyenne	Écart-type	Médiane	Medoid	Min	Max
bolus	23.06	11.43	21.00	32.00	8.00	43.00
basal	17.59	7.44	17.00	24.00	5.00	34.00
Poids	73.74	13.36	73.00	66.00	52.00	100.30
age	80.18	4.26	80.00	75.00	74.00	88.00
durée	25.24	11.13	22.00	37.00	10.00	42.00
imc	30.97	6.68	32.00	25.00	21.00	44.00
HBA1C	9.07	0.79	9.10	9.00	7.70	10.30
MDRD	63.53	27.02	58.93	83.40	22.95	134.46
doseTotal	40.65	16.74	34.00	56.00	13.00	70.00
ukg	0.56	0.26	0.45	0.84	0.24	1.11

Classe 2

Variable	Moyenne	Écart-type	Médiane	Medoid	Min	Max
bolus	56.40	27.90	54.50	44.00	22.00	123.00
basal	33.85	17.27	28.00	25.00	11.00	80.00
poids	86.53	13.45	90.65	93.40	52.10	106.00
age	69.45	4.21	68.50	65.00	65.00	79.00
duree	13.85	5.73	15.00	17.00	3.00	27.00
imc	31.94	4.53	32.75	34.00	22.00	40.00
HBA1C	8.97	1.56	8.75	7.90	6.50	12.50
MDRD	66.00	18.41	69.14	87.07	30.50	102.43
doseTotal	90.25	39.53	84.00	69.00	38.00	177.00
ukg	1.03	0.40	0.91	0.73	0.43	1.93

Classe 3

Variable	Moyenne	Écart-type	Médiane	Medoid	Min	Max
bolus	60.25	27.72	47.50	97.00	19.00	108.00
basal	44.19	20.47	35.50	36.00	28.00	93.00
poids	112.64	15.79	109.05	121.50	94.00	148.00
age	72.62	5.43	73.50	69.00	65.00	84.00
duree	24.88	10.76	23.50	32.00	10.00	42.00
imc	41.52	4.67	41.00	41.00	34.00	51.30
HBA1C	8.80	1.43	8.60	12.10	5.90	12.10
MDRD	50.95	19.73	47.33	32.00	15.60	83.40
doseTotal	104.44	43.64	83.00	133.00	51.00	192.00
ukg	0.94	0.44	0.78	1.09	0.52	1.83

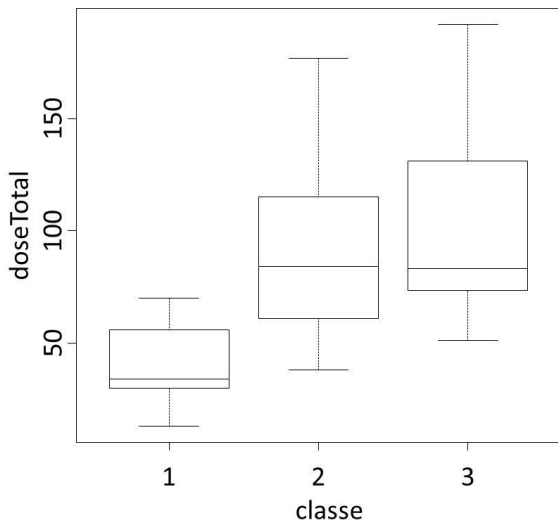


Fig. 4. Distribution de la dose totale d'insuline dans chacune des trois classes

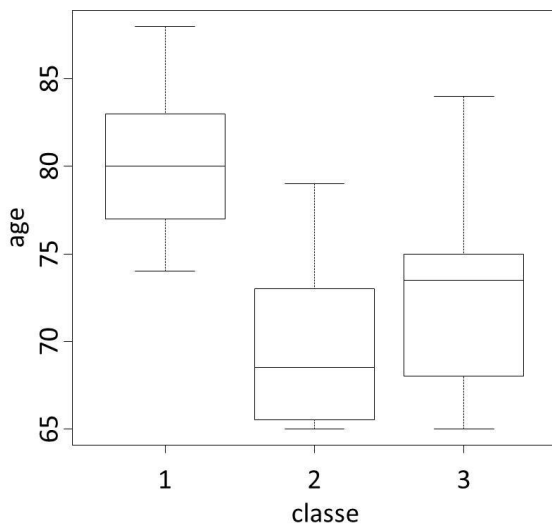


Fig. 5. Distribution de l'âge des patients dans chacune des trois classes

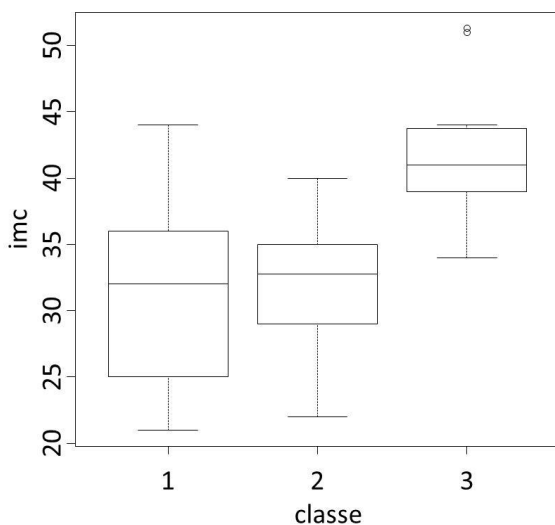


Fig. 6. Distribution de l'imc des patients dans chacune des trois classes

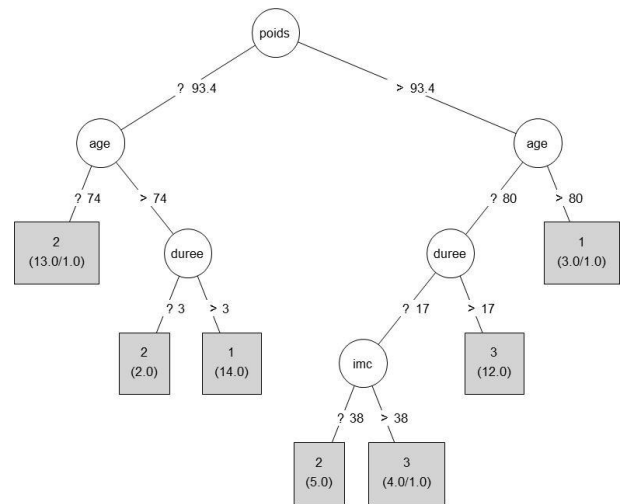


Fig. 7. Arbre de décision obtenue par l'algorithme C4.5 avec le logiciel R

La comparaison des distributions des valeurs des variables quantitatives dans les sous-échantillons définis par les classes, permet de guider l'interprétation en proposant des éléments de discrimination de ces classes. En observant la répartition de la dose insulínique totale prescrite (*doseTotal*) on observe tout d'abord que les distributions de ces doses, au sein des clusters, permettent de distinguer la Classe 1 des deux autres (figure 4).

Cette classe se caractérise donc, notamment par une dose insulínique totale faible. De la même manière les distributions de la variable *age* suggèrent que cette Classe 1 est constituée de patients plus âgés que dans les autres classes (figure 5).

La variable *imc* permet de différencier les classes 2 et 3 et de présenter la Classe 3 se distingue des deux autres par une *imc* plus élevée des patients qui la composent (figure 6).

Cette approche « interactive » de description des classes permet de proposer une première interprétation de ces regroupements.

C. Arbre de décision

L'utilisation d'algorithmes spécifiques permet de généraliser et d'automatiser cette étape d'explication. Les arbres de décision et autres méthodes discriminantes sont des outils particulièrement adaptés dans ce contexte.

Les arbres de décision [13, 16] présentent l'avantage de proposer des résultats généralement plus facilement interprétables que ceux des méthodes concurrentes. La figure 7 illustre le résultat obtenu à l'aide de l'algorithme C4.5 de Quinlin [16]. Le partitionnement récursif présenté permet de proposer une lecture explicite des classes 1, 2 et 3. Cet arbre de décision propose aussi des règles de classement des individus dans les classes interprétées. Dans des travaux médicaux précédents (en cours de relecture), il a été validé que l'âge des patients était un facteur discriminant de la variabilité des besoins insulíniques. L'extraction automatique proposé dans cette communication, par arbre de décision, confirme ce fait et permet de le situer parmi les autres variables potentiellement discriminantes.

IV. DISCUSSION ET CONCLUSION

Dans l'étape prédictive qui fera suite à ce travail, il s'avèrera plus efficace d'utiliser des algorithmes différents des arbres de décision. Comme nous l'avons dit avant, les arbres de décision proposent des modèles faciles à interpréter. Cette interprétabilité était recherchée dans le contexte explicatif de cet article. Les forêts aléatoires et les SVM sont plus difficilement lisibles mais donnent des résultats plus satisfaisants dans un contexte prédictif.

L'étude présentée dans cette communication confirme la complexité de la prescription insulinique chez le diabétique de type 2. La variabilité des besoins insuliniques est telle que la simple adaptation du schéma insulinique de type 1 n'est pas envisageable pour le diabétique âgé de type 2. Il est nécessaire de développer un système spécifique d'aide à l'insulinothérapie chez ces patients âgés diabétiques.

Dans ce contexte, l'étude montre trois classes de prescription de l'insuline correspondant à des doses différentes. Le schéma thérapeutique classique, pris par les diabétoles, consiste en une adaptation de la dose prescrite pour atteindre les objectifs glycémiques de chaque patient. L'utilisation de ces trois classes permet une adaptation plus systématique et plus rapide des doses prescrites. Ce résultat, appliqué en milieu hospitalier, devrait permettre de diminuer les temps d'hospitalisation. La technique proposée nécessite une étude de validation complémentaire.

Dans cette communication, nous n'avons pas présenté l'individualisation de l'objectif glycémique, particulièrement important dans le cas du diabète de type 2. Dans un modèle d'aide à l'insulinothérapie, l'évaluation de l'écart à cet objectif individuel est un élément crucial qui permet un meilleur contrôle du diabète et une diminution du risque de complications.

Lors d'un traitement insulinique, le risque principal est l'hypoglycémie. Ce risque est fonction de la stabilité du diabète et des objectifs d'équilibre glycémique individuel de chaque patient. Il dépend aussi de l'âge du patient, de son alimentation, de son activité physique et d'éventuelles pathologies intercurrentes. Pour réduire ce risque, notre système d'aide devra participer à une information et une formation du patient et de son entourage.

Dans le domaine de la télévigilance, le dispositif d'accompagnement thérapeutique sur lequel nous travaillons est complémentaire d'un dispositif d'alarme destiné à prévenir ou alerter en cas de d'hypoglycémie ou d'hyperglycémie. Un système global d'aide à l'insulinothérapie chez les personnes âgées doit intégrer les données de plusieurs capteurs pour évaluer les apports glucidiques, l'activité physique et les capacités cognitives et fonctionnelles. Un tel système doit aussi servir à la gestion de l'insulinothérapie dans sa dimension autre que physiologique.

Le modèle proposé dans cette communication concourt à l'émergence d'un système d'aide à la prescription insulinique chez les personnes âgées. Il est en cours de développement sur des terminaux portables mobiles et interactifs (tablettes tactiles) chez AXON' dans la région Champagne-Ardenne. Il sera utilisé dans des réseaux de soins comprenant généralistes,

spécialistes et centres diabétologiques pour une bonne prise en charge et un suivi régulier des patients diabétiques âgés de type 2.

REFERENCES

- [1] Bulletin épidémiologique hebdomadaire (BEH) de novembre 2010 (N° 42-43).
- [2] K.Joinville et al. Diabète traité en France en 2007 : un taux de prévalence proche de 4 % et des disparités géographiques croissantes. *Bull Epidémiol Hebd* 2008;43:409-413.
- [3] D.C.Klonoff. "using Telemedicine to Improve Outcomes in Diabetes— An Emerging Technology". PMID: PMC2769943.
- [4] R.Bellazi et al. "A telemedicine support for diabetes management". The T-IDDM project" *Computer Methods and Programs in Biomedicine* 69 (2002) 147–161.
- [5] O.k.Hejlesen et al. "DIAS-The diabetes advisory system: an outline of the system and the evaluation results obtained so far." *Comput Methods programs Biomed*.1997 sep; 54(1-2):49-58.
- [6] M.Cerossi et al. "Diabetes Interactive Diary (DID)". A new telemedicine system enabling flexible diet and insulin therapy while improving the quality of life: an open label, international, multicentre, randomized study ". *Diabetes Care* October 6, 2009.
- [7] D.Dardari et al." Téléphone portable et traitement des diabètes". *TH* 703, mai-juin 2007.
- [8] EJ. Gomez, M.Hernando, et al. "Telemedicine as a tool for intensive management of diabetes: the DIABTel experience". *ComputMethods Programs Biomed* 2002; 69: 163-77.
- [9] CA.Hilliman et al. "The Effects of Redesigning the IDEATel Architecture on Glucose Uploads". *Telemedicine Journal and e-Health*. 2009 April; 15(3): 248–254.
- [10] C.Quinn et al." Welldoc Mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction". *Diabetes Technology Ther* 2008 ; 10 : 160-8.
- [11] S.Franc et al." Télémedecine et diabète : état de l'art et perspectives". *Sang Thrombose Vaisseaux* 2011; 23, no 4: 178-86.
- [12] Sophia : " le service d'accompagnement de l'Assurance-maladie pour les malades chroniques". www.sophia-infoservice.fr
- [13] Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) *Classification and Regression Trees*. Wadsworth.
- [14] L. Kaufman et P.J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York, 1990.
- [15] A.K. Jain, M.N. Murty, P.J. Flynn. *Data Clustering: A Review*. *ACM Computing Surveys*, Vol. 31, No. 3, 1999.
- [16] Quinlan,J.R. *C4.5 : programs for Machine Learning*.Morgan Kaufmann Publishers,1993.