

# Apports de l'Intelligence Artificielle à la prédiction des durées de séjours hospitaliers.

Rachda Naila Mekhaldi<sup>1</sup>, Patrice Caulier<sup>1</sup>, Sondes Chaabane<sup>1</sup>, Sylvain Piechowiak<sup>1</sup>, Abdelahad Chraibi<sup>2</sup>

<sup>1</sup> LAMIH UMR CNRS 8201, Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielle et Humaine, Valenciennes, France  
{rachdanaila.mekhaldi, patrice.caulier, sondes.chaabane, sylvain.piechowiak}@uphf.fr

<sup>2</sup> ALICANTE, SECLIN, FRANCE  
abdelahad.chraibi@alicante.fr

## Résumé :

Au cours des dernières années, les services hospitaliers ont tenté d'optimiser leurs ressources afin d'améliorer le rendement du fonctionnement de l'hôpital. La durée de séjour (DDS) est l'un des indicateurs les plus importants pour évaluer ce rendement. Dans un premier temps, cet article identifie les facteurs qui affectent les DDS dans différents services hospitaliers. Un modèle de ces facteurs est alors proposé. Dans un second temps et afin d'analyser les facteurs du modèle et de prédire les DDS, l'article met en évidence un ensemble de techniques et d'applications d'intelligence artificielle dont le data mining et les algorithmes d'apprentissage automatique. Les Réseaux de Neurones Artificiels (RNA), les Arbres de Décision (AD) et les Vecteurs à Support Machine (VSM) sont souvent utilisés dans ce domaine. Finalement, une voie de solution est tracée afin d'améliorer la prédiction des DDS et, par conséquent, d'augmenter la qualité des soins, d'optimiser la gestion des ressources des hôpitaux et, enfin, de diminuer les coûts.

**Mots-clés** : Durées de séjours hospitaliers, exploration de données, apprentissage automatique, modèle de prédiction.

## 1 Introduction

Les établissements de soins cherchent sans cesse à optimiser le fonctionnement de leurs services tout en assurant la qualité de ces services. La planification des activités et la gestion des ressources ont un impact important sur cet objectif. L'estimation de la durée de séjour d'un patient au moment de son admission et durant son séjour est un indicateur d'évaluation de base des services de santé. La durée de séjour est définie comme l'intervalle de temps entre l'admission du patient et sa sortie (Khosravizadeh *et al.*, 2016). La prédiction des DDS hospitalier contribue à l'amélioration de la qualité des soins ainsi qu'à l'efficacité de la charge de travail opérationnelle. Cela permet une planification précise des réadmissions, une minimisation des coûts et une réduction du nombre de lits mal occupés. Afin de planifier les activités de soins de manière pertinente, des données de santé sous format électronique sont utilisées pour l'analyse de la variable DDS.

La DDS constitue un indicateur très important d'évaluation des performances des hôpitaux, plusieurs facteurs ont une incidence significative sur la DDS (Carter & Potts, 2014). L'objectif de ce papier est, d'une part, de recenser les différents modèles de DDS existants dans un environnement hospitalier puis d'en déduire un modèle généralisé. D'autre part ce papier vise à, mettre en place les méthodes et techniques d'intelligence artificielle dont le data mining et l'apprentissage automatique pour l'extraction des connaissances à partir des données médicales ainsi que la prédiction des durées de séjours. En effet, la disponibilité croissante de ces données contribue à améliorer les connaissances pour une meilleure prise en charge des patients et a élargi les champs de recherche dans la santé. Néanmoins, il existe des règles de confidentialité à respecter pour accéder à ces données. Le dossier d'un patient existe sous format électronique. Il est nommé DSE (Dossier de Santé Electronique). Chaque DSE est créé, géré et conservé par un organisme de santé et seuls les professionnels de santé qui participent aux soins d'un patient peuvent y accéder (Tom Seymour, Dean Frantsvog,

2012). Les données du DES concernent des informations générales, des diagnostics, les traitements du patient et son historique médical. D'autres formes de données peuvent porter sur des informations administratives, des modalités d'assurance, etc. Ces éléments sont hétérogènes et proviennent de multiples sources. Ce qui rend leur traitement complexe. De plus, les bases de données de santé sont souvent volumineuses et incomplètes.

L'objectif de l'étude est de présenter, différents modèles de durées de séjours hospitaliers existants dans la littérature puis de déduire un modèle généralisé. Ensuite, nous exposons les méthodes et techniques employées pour la prédiction de cette DDS et nous proposons une démarche méthodologique générique pour la prédiction des DDS dans un environnement hospitalier.

## 2 Facteurs influençant les DDS

De grandes quantités de données sur les séjours hospitaliers ont été stockées depuis plusieurs années. Pouvoir en extraire un modèle de prédiction des durées de séjour permettra de faire face aux nouvelles situations. Afin de construire ce modèle, il est indispensable d'identifier tous les éléments qui le constituent. La modélisation des durées de séjour en milieu hospitalier a pris différentes formes. En effet, des recherches antérieures ont tenté de grouper les patients en fonction de leur état de santé, en supposant que chaque maladie est associée à une durée de séjour recommandée (Shea *et al.*, 1995) par conséquent, la durée de séjour est définie différemment d'un service hospitalier à un autre. Dans (Music & Hindi, 2014), le type du service de l'hôpital est considéré comme un paramètre important pour la prédiction de la durée de séjour parmi les facteurs suivants : le nom de l'hôpital, le type du service, le nombre de lits libres, la politique de sortie. Nous avons étudié les différents modèles de durées de séjour hospitalier selon un type de service bien défini. Nous avons constaté que les auteurs se focalisent souvent sur les services hospitaliers suivants : le service de chirurgie générale, le service d'urgence, le service de cardiologie et le service de soins intensifs.

Dans ce qui suit, nous présentons les facteurs qui influencent les DDS par type de service. Divers modèles de DDS sont donc mis en œuvre. Un modèle générique de DDS est enfin donné.

### 2.1 Facteurs influençant les DDS dans un service de cardiologie

Plusieurs travaux s'intéressent à la prédiction des DDS dans les services de cardiologie. Dans l'étude menée par (Lafaro *et al.*, 2015), 8 variables ont été sélectionnées parmi 36 pour concevoir le modèle de prédiction des DDS. Ces variables liées à l'unité de soins intensifs chirurgicaux cardiaques sont : l'utilisation d'une pompe de ballon intra-aortique, niveau d'émission de l'O<sub>2</sub>, utilisation d'agents isotropes cardiaques positifs, l'hématocrite, le taux de sérum créatinine et la mesure d'analyse du gaz du sang. Les facteurs utilisés dans (Hachesu *et al.*, 2013) sont la consommation des médicaments anticoagulants et de la nitrate, mesure de la pression artérielle diastolique et de cholestérol, douleur thoracique, fraction d'éjection, comorbidité associée, densité de lipoprotéine, si le patient est fumeur ou non et le taux d'hémoglobine. Dans l'étude de (Tsai *et al.*, 2016), l'adresse du patient, son diagnostic, le type de l'intervention, la comorbidité et le mode de remboursement permettront de construire le modèle de la prédiction des DDS. Dans (Almashrafi *et al.*, 2016) les auteurs ont analysé les éléments qui influencent une longue durée de séjour dans le service de soins intensifs cardiaque. Ils ont cité l'historique des maladies du patient, le nombre de complications, la fraction d'éjection ventriculaire gauche, etc. Ces éléments sont disponibles au moment de l'admission du patient. Pour l'ensemble de ces études, l'âge et le sexe du patient sont toujours utilisés. A partir des éléments décrits ci-dessus, nous avons constaté que la modélisation de la durée de séjour dans un service de cardiologie demande des informations qui ne sont pas disponibles au moment de l'admission du patient. Comme nous avons également constaté qu'il existe plusieurs facteurs communs dans les différentes études. Essentiellement, les facteurs démographiques, la mesure du taux de créatinine, le taux de l'hématocrite, etc. Ces facteurs sont représentés avec des termes différents dans la littérature mais ils portent la

même signification. De ce fait, l'aide des experts dans le domaine médical est primordial pour avoir un ensemble d'attributs cohérent pour caractériser les DDS dans la phase de sélection de variables.

## 2.2 Facteurs influençant les DDS dans un service de soins intensifs (SSI)

Plusieurs travaux se sont intéressés à l'analyse des facteurs qui influencent les DDS dans les unités de soins intensifs en général. La base de données MIMIC III est utilisée pour classer les DDS en deux groupes : longue et courte après avoir quitté la SSI. Les facteurs utilisés concernent l'admission, le transfert, la sortie du patient, les prescriptions médicales, les tests du laboratoire, le diagnostic et les facteurs démographiques (Gentimis *et al.*, 2017). Les données concernant 311 patients ont été utilisées dans (Maharlou *et al.*, 2018) pour identifier les facteurs qui impactent les DDS. Selon cette étude, les facteurs démographiques, l'historique médical (maladie cardiaque, rénale, maladie pulmonaire), taux de créatinine, rythme cardiaque, etc constituent ces attributs. Nous avons remarqué que le modèle de DDS dans un SSI est complexe et nécessite une bonne connaissance du domaine de la santé.

## 2.3 Facteurs influençant les DDS dans un service de chirurgie

Différentes études ont ciblé le service de chirurgie pour analyser les variables utilisées caractéristiques des DDS. La prédiction des DDS est utile en phase de préopération ou postopération. Dans (Chuang *et al.*, 2016), les auteurs ont montré que dans un même service de chirurgie, les facteurs sont différents pour une opération urgente et pour une opération non urgente. Ces facteurs sont : les informations démographiques, l'historique médical du patient, les mesures des signes vitaux et des résultats du laboratoire, et les données sur l'opération reportées par le médecin et les infirmiers. Dans (Khosravizadeh *et al.*, 2016), il a été montré que les variables qui influencent le plus la DDS sont : la situation familiale du patient, ses conditions de sortie, le type du traitement ainsi que le type de paiement pour les dépenses de l'hôpital. Les recherches dans (Aghajani & Kargari, 2016) se sont focalisées sur la sélection des variables suivantes : le type de l'opération, le nombre d'opérations qu'un patient a subi, le temps entre l'ordre de sortie et la sortie effective du patient, les informations de transfert entre services, le nombre moyens de visites par jour, le nombre de consultations médicales, les hospitalisations antécédentes du patient et le nombre de tests effectués pour le service de chirurgie générale.

## 3 Analyse et discussion à propos des facteurs influençant les DDS

Suite à ces recherches, nous avons analysé ces différents facteurs et nous avons conclu ce qui suit :

- La prédiction des durées de séjour hospitalier est déterminée par le type de service étudié.
- les services de cardiologie sont souvent les plus étudiés. Cela peut être dû au fait que ce service demande une plus grande part de financement de l'hôpital.
- Il existe des facteurs en communs avec les services de cardiologie, de soins intensifs et de chirurgie. Nous retrouvons l'historique médical du patient, son rythme cardiaque, analyses du laboratoire, etc. Les facteurs démographiques tels que l'âge, le genre et la situation familiale du patient sont souvent utilisés.
- Pour un modèle plus généralisé, il est essentiel de se référer aux spécialistes dans le domaine médical afin de nous guider dans notre sélection de variables comme une première étape de notre processus d'étude. Nous avons résumé les facteurs communs qui figurent dans les différents services dans le tableau 1.

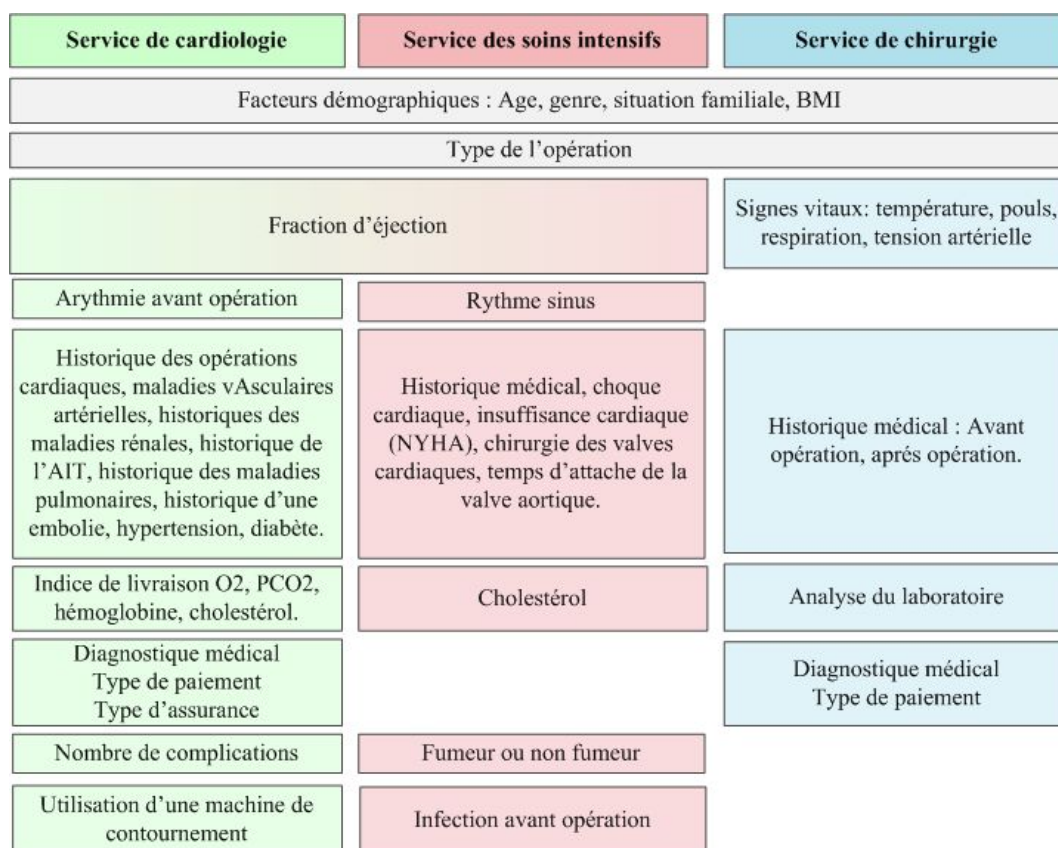


FIGURE 1 – Facteurs impactant les DDS.

Afin de sélectionner les facteurs qui influencent les DDS et de concevoir un modèle de prédiction des durées de séjour hospitalier, des techniques d'intelligence artificielle ont été utilisées. Dans la section suivante, nous présentons ces techniques et proposons par la suite notre solution.

#### 4 Techniques d'Intelligence Artificielle pour la prédiction des DDS

L'intelligence artificielle a fait un retour fracassant dans le domaine de la santé. Plusieurs algorithmes d'aide à la décision sont utilisés pour le diagnostic médical, l'analyse des images de radiologie, l'étude des parcours de patients, etc. Le choix des algorithmes d'apprentissage automatique dépend essentiellement de l'objectif de l'étude, de la taille, la qualité et la nature de la base de données. Dans (Liao *et al.*, 2016), les auteurs ont appliqué le clustering sur un ensemble de patients atteints d'une maladie rénale en phase terminale initiés à l'hémodialyse (HD). Ce groupement est basé sur un modèle de changement des coûts des soins de santé avant et après l'HD selon le profil des patients. Ce modèle utilise deux algorithmes K-means et CAH. Dans une autre étude, l'algorithme K-means est employé pour analyser la variation dans les volumes d'entrées et de sorties des hôpitaux et les regrouper selon l'affectation des ressources et l'efficacité des services de soins (Tseng *et al.*, 2015). Un intérêt important est donné aux méthodes du data mining pour l'analyse des données médicales et d'apprentissage automatique pour la classification de ces données. Les données médicales ont été largement utilisées ces dernières années dans le développement des systèmes de soins. Dans le volet de la prédiction des durées de séjour hospitalier, plusieurs travaux existent. Dans (Hachesu *et al.*, 2013), les auteurs ont comparé entre différents modèles de réseaux de neurones artificiels pour prédire la DDS dans l'unité de soins intensifs cardiaque. Une comparaison est

établie dans (Shea *et al.*, 1995) entre les réseaux de neurones et les réseaux de neurones à base de système d'inférence flous. En plus des réseaux de neurones, les forêts d'arbres décisionnelles sont implémentées dans (Pendharkar & Khurana, 2014). Les réseaux de neurones, les arbres de décision et les machines à vecteurs de support sont utilisés pour la prédiction des durées de séjours des patients souffrants d'une coronaropathie (Khosravizadeh *et al.*, 2016). Un modèle de prédiction des longues durées de séjour avant une opération à base des arbres de décision, SVM et le random forest a abouti à un meilleur résultat avec le random forest (Al-mashrafi *et al.*, 2016). D'autres méthodes d'apprentissage supervisé sont citées dans (Carter & Potts, 2014) telles que les arbres de régression et de classification (CART) et Chi-Squared Automatic Interaction Detector (CHAID). Les techniques de régression linéaire, l'algorithme naïf de Bayes et les k plus proches voisins sont appliquées pour déterminer les facteurs qui influencent les DDS dans un service de chirurgie générale (Gentimis *et al.*, 2017). Plusieurs types d'arbres de décision et de régression sont exploités dans l'étude (Tom Seymour, Dean Frantsvog, 2012) afin de décider si une DDS pour un patient après opération est dans un intervalle standardisé et dans (Tsai *et al.*, 2016) pour détecter des situations critiques dans le service d'urgence de pédiatrie. Les réseaux de neurones sont aussi utilisés dans (Lafaro *et al.*, 2015) pour la prédiction des durées de séjour pour les patients atteints de l'une de ces trois maladies : maladie de l'artère coronaire, arrêt cardiaque et infarctus aigu des myocards.

A partir de ces différentes recherches, nous constatons que les méthodes d'apprentissage supervisé sont employées pour le problème de prédiction. Il s'agit d'attribuer un séjour hospitalier à une catégorie définie. Les comparaisons entre les différentes méthodes citées auparavant dépend de l'étude. Généralement les algorithmes random forest et réseaux de neurones ont les meilleures précisions. Quoique ces méthodes aboutissent à des résultats satisfaisants, leur faiblesse réside dans le fait que ces méthodes sont utilisées pendant le séjour du patient et non pas juste au moment de son admission.

Pour pallier cette faiblesse, nous proposons une approche prédiction dynamique de la durée de séjour selon des intervalles de temps séparés. Pour un service donné, une première étape est de définir une durée minimale de séjour représentant le nombre de jours que le patient doit passer dans ce service. Une étape de sélection de variables est réalisée. Des méthodes de data mining et du clustering seront explorées pour restreindre l'ensemble de variables représentant les DDS. L'intervention des experts dans le domaine médical est nécessaire pour valider les résultats. Pour la phase de classification automatique, un premier modèle de prédiction est réalisé pendant le séjour une patient qui est inférieur à la durée minimale de séjour fixée auparavant puis enrichi au fur et à mesure pendant le séjour du patient. Le choix des intervalles de temps et les catégories des durées de séjour dans l'hôpital dépendront des expérimentations. Le diagramme de la figure 2 illustre cette approche.

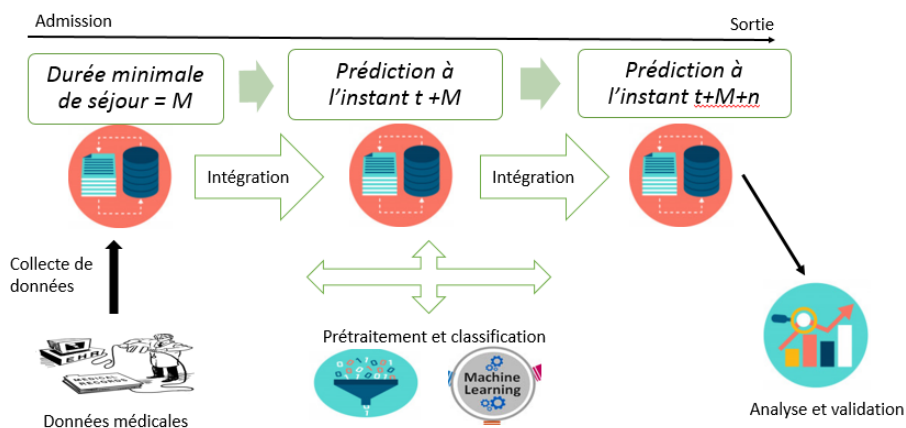


FIGURE 2 – Prédiction des DDS : proposition.

Pour mettre en évidence notre solution proposée, des expérimentations vont être menées

sur les données médicales du Centre Hospitalier de Valenciennes (CHV). Ces données sont de type PMSI (Programme de Médicalisation de Systèmes d'Information) et qui sont implémentés dans l'ensemble des établissements de soins français. Des informations disponibles pour l'ensemble des patients hospitalisés dans les services hospitaliers étudiés sont aussi intégrées aux données PMSI.

## 5 Conclusion

La prédiction des durées de séjour prend de l'ampleur ces dernières années afin de planifier les activités de soins de manière pertinente. C'est un indicateur de performance de la qualité de l'efficacité des services des hôpitaux. Dans ce travail, nous avons présenté l'utilité de la prédiction des durées de séjour hospitalier en donnant plusieurs modèles de durées de séjour. Cette problématique est complexe car elle dépend de nombreux facteurs. Nous avons mis en évidence les facteurs qui influencent la variable DDS. En se basant sur les principaux travaux de recherche effectués dans ce domaine. Ainsi, des solutions se basant sur les méthodes d'intelligence artificielle dont le data mining et l'apprentissage automatique sont exposées. Enfin, une démarche méthodologique générique est proposée pour la prédiction des durées de séjour dans un environnement hospitalier.

## Références

- AGHAJANI S. & KARGARI M. (2016). Determining Factors Influencing Length of Stay and Predicting Length of Stay Using Data Mining in the General Surgery Department. *Hospital practice and research (HPR)*, **1**(2), 53–58.
- ALMASHRAFI A., ALSABTI H., MUKADDIROV M., BALAN B. & AYLIN P. (2016). Factors associated with prolonged length of stay following cardiac surgery in a major referral hospital in Oman : A retrospective observational study. *BMJ Open*, **6**(6).
- CARTER E. M. & POTTS H. W. (2014). Predicting length of stay from an electronic patient record system : A primary total knee replacement example. *BMC Medical Informatics and Decision Making*, **14**(1), 1–13.
- CHUANG M. T., HU Y. H., TSAI C. F., LO C. L. & LIN W. C. (2016). The Identification of Prolonged Length of Stay for Surgery Patients. In *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, p. 3000–3003.
- GENTIMIS T., ALNASER A. J., DURANTE A., COOK K. & STEELE R. (2017). Predicting Hospital Length of Stay using Neural Networks on MIMIC III Data. In *IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, p. 1194–1201.
- HACHESU P. R., AHMADI M., ALIZADEH S. & SADOUGHI F. (2013). Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients. *Healthcare Informatics Research*, **19**(2), 121–129.
- KHOSRAVIZADEH O., VATANKHAH S., BASTANI P., KALHOR R., ALIREZAEI S. & DOOSTY F. (2016). Factors affecting length of stay in teaching hospitals of a middle-income country. *Electronic physician*, **8**(10), 3042–3047.
- LAFARO R. J., POTHULA S., KUBAL K. P., INCHIOSA M. A., POTHULA V. M., YUAN S. C., MAERZ D. A., MONTES L., OLESZKIEWICZ S. M., YUSUPOV A., PERLINE R. & INCHIOSA M. A. (2015). Neural Network Prediction of ICU Length of Stay Following Cardiac Surgery Based on Pre- Incision Variables. *plos one*, p. 1–19.
- LIAO M., LI Y., KIANIFARD F., OBI E. & ARCONA S. (2016). Cluster analysis and its application to healthcare claims data : A study of end-stage renal disease patients who initiated hemodialysis Epidemiology and Health Outcomes. *BMC Nephrology*, **17**(1), 1–14.
- MAHARLOU H., NIAKAN KALHORI S. R., SHAHBAZI S. & RAVANGARD R. (2018). Predicting length of stay in intensive care units after cardiac surgery : Comparison of artificial neural networks and adaptive neuro-fuzzy system. *Healthcare Informatics Research*, **24**(2), 109–117.
- MUSIC B. & HINDI Q. (2014). Machine learning techniques for predicting hospital length of stay in pennsylvania federal and specialty hospitals LENGTH OF STAY IN PENNSYLVANIA FEDERAL AND SPECIALTY. *International Journal of Computer Science and Applications*, **11**(3), 45–56.

- PENDHARKAR P. C. & KHURANA H. (2014). Machine learning techniques for predicting hospital length of stay in pennsylvania federal and specialty hospitals. *International Journal of Computer Science and Applications*.
- SHEA S., SIDELI R. V., DUMOUCHEL W., PULVER G., ARONS R. R. & CLAYTON P. D. (1995). Computer-generated informational messages directed to physicians : Effect on length of hospital stay. *Journal of the American Medical Informatics Association*, **2**(1), 58–64.
- TOM SEYMOUR, DEAN FRANTZVOG T. G. (2012). Electronic health records (EHR). *American Journal of health Sciences*, **3**(3).
- TSAI P. F. J., CHEN P. C., CHEN Y. Y., SONG H. Y., LIN H. M., LIN F. M. & HUANG Q. P. (2016). Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network. *Journal of Healthcare Engineering*, **2016**, 11 pages.
- TSENG S. F., LEE T. S. & DENG C. Y. (2015). Cluster analysis of medical service resources at district hospitals in Taiwan, 2007–2011. *Journal of the Chinese Medical Association*, **78**(12), 732–745.