International Arab Conference on Information Technology (ACIT'2016)

Decision Support System for identification and classification the Whewellite- and Weddellite crystals in human urinary using data minig algorithms

A. Ait Ider[1], D. Naji[1] C.Tcheka[2], A. Ben Ali[3], A. Merbouha[1] and M. Mbarki[1]

[1] University of Sultan Moulay Slimane, Faculty of Science and Technology, P.B 523, BeniMellal, Morocco
[2] University of Yaoundé 1, Faculty of Science, P.B 812, Yaounde, Cameroun.
[3] National Control Laboratory Medicines,Ministry of Health, Madinat Al Irfane, P.B 6206, Rabat, Morocco

Abstract: *The majority of the analyzed calculi from patients is composed of calcium oxalate (CaOx) monohydrate (whewellite (Wh)) and CaOx dihydrate (wedellite (Wd)). The urinary calculi were identified by chemical and morphological analysis based on106 urine samples from human voluntary. The Crystalluria made by an optical polarized light microscopy. The apparent oxaluria and urinary calcium were determined by conventional volumetric assays. The database contains 63 men and 43 women among whom 30 (28.3%) cases had a type of crystal CaOx (Wh), 27 (25.5%) cases had a type of crystal CaOx dihydrate (Wd) and 49 (46.2%) cases are not affected by this disease (Nc). The aim of this paper was to compare the performances of different algorithms techniques: Rnd Tree (Random Forest), C4.5 and Artificial Neural Networks (ANNs) algorithms in order to develop a simple system to predict and classify urinary calculi. The results obtained show that the Rnd Tree algorithm is a very powerful model and was superior to different techniques in prediction, the average correct classification rate achieved 97.39%. The accuracy of the system was determined by comparing the correct classification rates of the different algorithms.*

Keywords: *calcium oxalate; urinary stones; data mining algorithms; Intelligent Diagnosis System.*

1. Introduction

The Lithogenesis includes all the processes driving the development of stone in the urinary tract. Lithogenesis includes several phases which are expressed successively simultaneously[1]. The main and major factor resulting in crystal formation is supersaturation as the primary event in lithogenesis that is a process during which, the oversaturation of urines by the lithogenic substances behind the formation of crystals. These crystals can in certain conditions and according to the oversaturation degree, agglomerate may lead to the formation of urinary stones in the urinary system generating renal failure in human. In fact the urine composition (inlithogenic substances) has an important role and driving force in stone development. In fact the prevailing tendency of urolithiasis is increasing across the world [2, 3]. It is caused by several factors that include frequently CaOx crystalluria in the urine. It crystallizes in the form of (Wd) and (Wh) [2, 4- 8]. The previous studies have been indicated the variations in Ox concentration induce greater and important effects on CaOx supersaturation and in the number and size of CaOx crystals.Also have been indicated that the risk of forming CaOx crystals depends on both Ca and Ox concentrations, namely on the CaOx molar product, whereas the crystalline phase of CaOx, either mono or di-hydrate, depends on the molar ratio of Ca to Ox [9]. During diagnosis of the disease, no precise decision can be easily made taking into account the individual concentrations of calcium and oxalates those individually. While it is necessary to introduce the two important variables CaOx molar product (pCaOx) and the molar ratio calcium / oxalate (Ca/Ox) in order to ensure a better diagnosis [10]. It appears from the results of most studies in the literature that CaOx is the most frequent and majority urinary crystals crystallize [11- 16]. Previous studies

have shown that the combined consideration of Ca and Ox molar products and ratios may afford a better prediction of both the probability of forming CaOx crystals and the type of CaOx crystals than does the separate consideration of urinary Ca and Ox concentrations [9]. However, rarely are the studies dedicated to the investigation of the influence of the calciuria and the oxaluria on the formation of the crystals CaOx in urines. As a matter of fact, the crystals are a promoting phase to the formation of urinary stones. The work of Daudon et al. in 1987 on the study of oxalo-calcium crystalluria nephrolithiasis of subjects have shown the influence of the product of the concentrations molar urinary calciums and oxalates (pCa.Ox) and their ratio (Ca / Ox) on the crystallization of CaOx. Daudon et al. in 2004 confirm that the motor of the crystallization of CaOx is the molar oxalocalcique product, but the crystalline form in which it is observed in the urine is highly dependent on the ratio molar calcium / oxalate [18). It has been shown that the main determinant of the crystallization was the molar product oxalocalcique (pCaOx) Daudon et al. in 1989, also that crystal form whewellite or weddellite depends mainly (but not exclusively) of molar (calcium / oxalate (rCa / Ox) Daudon et al. in 2015. The observations, values Crystalluria and oxaluria are confirmed by those of Daudon et al in 1987. The results obtained in our laboratory [21] showed that there is a strong correlation between CaOx crystalluria and the urinary concentrations of calcium and oxalate. calciuries and oxaluries values even approximate, may allow to predict the formation of oxalate crystals calcium. The influence of the calcium concentrations and oxalate (Ca. and Ox) was valued. The presence of cristalluries increases with the concentrations of calcium and oxalate, on the one hand and with their products (pCa.Ox.) and ratio (Ca / Ox.), on the other hand. In the present work, different classification algorithms have been applied, in order to classify the subjects of the database into 3 classes, depending on the chemical concentrations of variables (Ca, Ox, pCa.Ox, and Ca / Ox) and on their clinical status. The first class contains subjects presenting a Wh crystal type. The second class contains subjects with a Wd crystal type and the third one includes subjects who are unaffected by this disease According to the obtained results of the classification, we developed a simple intelligent diagnosis system. This system is important for medical laboratories since it can provide useful information about the development, formation of the urinary stones. The determination of the crystal types of the urinary stones reddwould be helpful to better manage, economically and socially, the treatment of this rising disease [16].

2. The Research Method

2.1 Datasets

Morphological analysis

The first phase of the analysis is the determination of the morphological type of crystals according to the classification (Daudon M. et al., 1999). The morphological analysis was conducted in accordance with the published protocol. The study was carried out on 106 urine samples from human volunteers from the region of Tadla Azilal.; all samples were analyzed and prepared in our laboratory [21]. Crystalluria has been characterized using optical polarized light microscopy.

Chemical analysis

The apparent oxaluria and urinary calcium were determined by conventional volumetric assays on urines. The database contains 63 men and 43 women among whom 30 (28.3%) cases presented (Wh) crystal type their urine, 27 (25.5%) cases presented a (wd) crystal type and 49 (46.2%) cases would be not affected by this disease (Nc)

2.2 Data analysis

Random forest

The random forests proposed by (Breiman 2001), which add an additional layer of randomness to bagging. In addition to constructing each tree using a different bootstrap sample of the data, random forests change how the classification or regression trees are constructed. The Rnd Tree algorithm generates multiple trees without pruning. In the training procedure, each tree is built based on a random subset of the original data (with replacement). In addition, a randomly selected subset of predictors is chosen for each built tree (Breiman, 2001).

C4.5 Algorithm

C4.5 is a supervised learning classification algorithm used to construct decision trees from the data (Quinlan, 1993). A C4.5 decision tree is one of the most widely used inductive inference tools [23]. The tree is generally constructed in a top down manner. The construction begins at the root node where each attribute is evaluated using a statistical test to determine how well it can classify the training samples. The best attribute is chosen as the test at the root node of the tree.

Algorithm of Artificial Neural Networks

The Artificial Neural Networks (ANNs) are used in various disciplines such as economics, ecology, environment, biology and medicine. These networks are applied in particular to solve problems of classification, prediction [24-26], categorization, optimization, recognition of the forms and the associative memory [27]. The ANN improves the diagnosis in medical sciences. The ANN was applied in order to determine the chemical composition of

urinary stones and to classify them [28]. In fact, the ANNs are used, for example, in cases of the prediction of myocardial [29, 30], lung pathologies [31, 32], diabetes [33], cancers [34], Alzheimer's disease [35] electromyography and kinesiology [36]. In an ANN, the Neurons are connected via synapses (connection links), that modulate signals passing through them. Each synapse has an associated weight w. The net input N is the function of all transmitted signals Xi and their corresponding weights W_i in a neuron: $N=\sum (W_i X_i)$ (sum of weighted input signals). Each neuron applies an activation (transfer) function to its net input N in order to provide an output signal for each neuron. A neural network is characterized by its architecture or its pattern of connections between the neurons. Several research works were made to describe the functioning of ANNs [37-39].

3. Results and Discussion

The determination of calciuria and oxaluria, even approximate, can allow predicting the CaOx crystal formation in the lithogenic patients urines. The differentiation of crystalline phases of cristalluries is important in the search for causes crystal growth or even lithogenesis. Otherwise, if information is disclosed too late it will be difficult to assess the usefulness and relevance of the information. The study of the spontaneous crystalluria is source essential information for the etiological diagnosis and medical management of patients with urinary stones or crystal pathologies likely have deleterious consequences for the function renal. It should be practiced in all laboratories to enable better detection of risk factors and more effective monitoring of patients suffering from urolithiasis [18]. In this study two main objectives of this work: the first objective was the search of the offending parameters in the development of urolithiasis. It should be noted, in agreement with previous work that the calculi remain the irreplaceable source of information on his training and analysis by appropriate physical techniques leading to the knowledge of its chemical composition allows access to the etiology of urolithiasis. Also, crystal growth is the decisive step of lithogenesis in which the crystallizable certain substances in urine. Therefore, the study of crystalluria is crucial in understanding the mechanisms that govern the initial phase of development of urolithiasis. The second objective was to study the influence of CaOx concentrations on the formation of urinary crystals as well as to classify and identify of urinary stone based on these parameters using data mining algorithms. Several classification algorithms were applied to the same database in order to develop a simple intelligent diagnosis system comparing the model performances of these supervised algorithms. the results of comparing the performance of ANN algorithms, k-nearest neighbor (KNN) - and Support Vector Machines (SVM) in the previous study showed that ANN are very powerful methods to develop a simple diagnostic system [39]. In this study the obtained results show that Rnd Tree algorithm is very powerful methods in prediction. The best results are obtained with the Rnd Tree technique, with a lower error rate than those obtained in the case of the other classification methods. The complexity of the neural network is measured by the number of neurons in the hidden layer. The number of hidden layers and the number of neurons from layer have an influence on the quality of learning. The choice of number of hidden layers is empirical. The comparison of the performances of all the models is presented in table 3 and figure 3. It is clear that the error rate generated by Rnd Tree Algorithm is very less compared to all other Algorithms. Using KNN, the error classification rate for the test samples is: 0.3445 for Wh, 0.2275 for Wd and 0.3015 for Nc. Using SVM, the error classification rate is 0.174 for Wh, 0.058 for Wd and 0.12 for Nc. The error prediction rate, using ANN, is 0.165 for Whewellite, 0.066 for Weddellite and 0.083 for Nc. The error classification rate is 0.25 for Wh, for Wd 0.44, for Nc 0.323 using C.4 and 0.784 for Nc using Rnd Tree (see table 2 and figure 1). It should denote that in Spite of Our technique for studying the influence of urinary calcium and oxaluria on the crystallization whewellite and weddellite proved preliminary and limited in terms of the approximate chemical assay techniques the best average correct prediction achieved was 97.39 % using Rnd Tree. Using ANN, SVM, KNN and C4.5 the average correct prediction was achieved 69.85%, 88.2%, 89.53% and 67.70 respectively (see table 3 and figure 2).

Therefore, it turns out that the average correct prediction rate of Rnd Tree is higher than the one obtained by using SVM, ANN, C.4 and KNN techniques. However the error rate of Rnd Tree in classification of class Nc was superior to the other algorithms applied in this study. The Rnd Tree algorithm is not mistaken for those of class 3 and cannot correctly classify some samples. The frequency of crystalluria oxalocalcique increases proportionally with the oxalocalcique molar product (pCaOx) both in the lithiasic patient than in normal subjects [18]. The performance of Rnd Tree is presented in Table 1 using a confusion Matrix. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. Using a confusion matrix it is easy to see if the system is confusing tree classes of urinary stones.

The obtained results show that the Rnd Tree is a helpful tool for the identification these types of stone. The Chemometrics coupled to simpler and less expensive chemical and physical analyses (volumetry

and optical microscopy), on the one hand and methods of data analysis, on the other hand would be a quick and less expensive tool for a first diagnosis of the urolithiasis that can allow the decision-makers to avoid the high budgets, of coverage, assigned to the treatments of this disease. Thanks to Rnd Tree technique, we have been able to establish a decision-making model allowing the competent sanitary services to predict the lithogenous risk in the populations. The study of crystalluria is an excellent review of laboratory that brings many benefits to the diagnosis and management of kidney diseases crystalline origin [41]. The results of most works in the literature showed that CaOxs is the most frequent as chemical species in urine and the majority of stones are crystalluria [10- 15]. However, rarely are the studies dedicated to the investigation of the influence of urinary calcium and apparent oxaluria on the formation of CaOx crystals in the urine; that is the initial step on the formation of urinary stones. These crystals can in some conditions become stones that lead to the failure of the urinary system.

Table 1. Confusion matrix using Random Forest

Confusion matrix				
	Wh	**wd**	**Nc**	**Sum**
Wh	30	0	0	30
Wd	0	27	0	27
Nc	3	1	45	49
Sum	33	28	45	106

Table 2. Comparison of the Errors Rate

Algorithms	**Wd**	**Wh**	**Nc**
RND Tree	0.00	0,00	0.78
C4.5	0.44	0.25	0.32
kNN	0.23	0.34	0,70
SVM	0.06	0.17	0,12
ANN	0.07	0.17	0.08

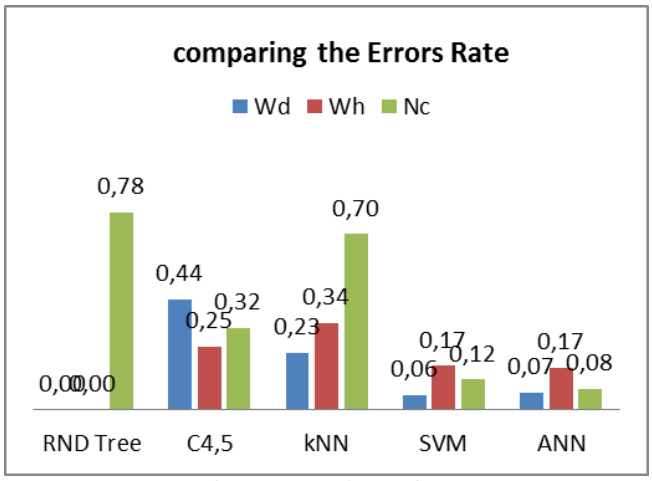

Figure 1. Comparison of error rates of classification algorithms

Table 3. Comparison of Correct classification rate

Algorithms	**Wd**	**Wh**	**Nc**	**Correct classification rate**
RND Tree	100.00%	100.00%	92.16%	97.39%
C4,5	75.00%	55.56%	73%	67.70%
kNN	65.55%	77.25%	69.77%	69.85%
SVM	82.60%	94.20%	87.80%	88.20%
ANN	83.50%	93.40%	91.70%	89.53%

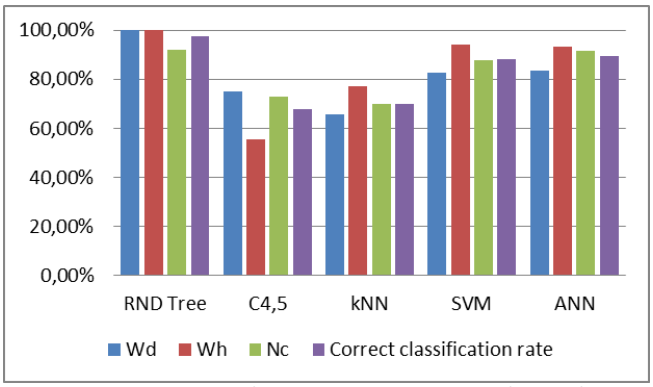

Figure. 2. Comparison of correct prediction rate of classification algorithms

4. Conclusion

This illustrates that the use of a simple and non-expensive physicochemical analyses coupled to classification algorithms have been successfully employed. This coupling technique is fast and simple. It can be very useful for the first diagnostic which is of great advantage in this context. Based on the results the Rnd Tree Algorithm produced less error rates when compared to all other Algorithms. This work has shown that Rnd Tree is very powerful method to develop a simple diagnostic system since it achieved a good predictive performance and yields the highest average correct prediction rate. According to the

obtained results, the Rnd Tree was recommended to develop a helpful decision support system.

Conflict of interest

The authors declare that they have no conflict of interest

References

[1] M. Daudon, O. Traxer, E. Lechevallier, C. Saussine (2008) La lithogenèse. Progrès en urologie 18, 815—827

[2] Renard-Penat R, Ayed A. (2010) Diagnostic et bilan des calculs urinaires.Radiologie et imagerie médicale : Génito-urinaire - Gynéco-obstétricale – Mammaire.334-173-C-10.

[3] .Fekak H, Sahnoun A, Rabii R et coll. (2006) La lithiase urinaire : étude spectrophotométrique infrarouge (à propos de 80 cas). J Maroc Urol;1:17-20.

[4] Daudon M, Traxer O, Lechevallier E, Saussine C. (2008) La lithogenèse. Prog Urol ; 18 : 815 27.

[5] Daudon M, Hennequin, Jungers P, Lacour B, Drueke T. (1994) Les inhibiteurs de cristallisation. In: Funck-Brentano JL, Bach JF, Kreis H,Grunfeld JP (éds). Actualités Néphrologiqes Jean Hamburger - Hôpital Necker, Paris, Médecine-sciences Flammarion 165-22

[6] Marangella M. Daniele P G, Ronzani M, Sonego S, Linari F. (1985) Urine saturation with calcium salts in normal subjects and idiopathic calcium stone-formers estimated by an improved computer model system Urol Res 189-193.

[7] Daudon M. (2004) Les différents types de calculs: apport de l'analyse morphoconstitutionnelle au diagnostic étiologique de la lithiase. In: Doré B (ed) Les Lithiases Rénales. Paris, Springer, pp.49-77

[8] Jungers P, Daudon M, Le Duc A. (1989) Paris Flammarion Médecine-Sciences.

[9] M. Daudon R. J. Réveillaud , H. Levasseur , P. Jungers. (1989) Combined Influence of Urinary Calcium and Oxalate Concentrations on Crystal Formation in Stone Formers. pp 155-157 Springer US doi: 10.1007/978-1-4899-0873-5_43

[10] Jungers P, Daudon M, Conort P. (1999) Lithiase rénale: diagnostic et traitement, ed Flammarion, paris,.

[11] Dao NQ, Daudon M. (1997) Infrared, and Raman spectra of stones. Paris: Elsevier,.

[12] Graf JD. Féraille E. (2000) Importance de l'analyse du calcul dans l'exploration de la lithiase rénale. Méd Hyg; 58: 431-433.

[13] Laziri F, Rhazi Filali F, Oussama A, Soulaymani A, Qarro A, Lezrek M. (2010) Facteurs impliqués dans l'épidémiologie des calculs urinaires marocains. J Maroc Urol ; 19 : 9 14.

[14] Belhaj R, Hayder N, Gargueh T, Zorguati M, Marrakchi O, Abdelhak S, Lakhoua R, Abdelmoula J. Diagnostic biochimique et moléculaire de l'hyperoxalurie primaire de type 1 : étude tunisienne à propos de 15 cas J. Pathologie Biologie 59 (4), pp. e97-e102

[15] Amrani Hassani M, Hennequin C, Lacour B, Daudon M. (2005) Citraturie et cristallurie de weddellite. Prog Urol 15: 650-5.

[16] Monnier L, Piperno M, Colette C, J.-L. Schlienger: Applying nutritional and metabolic perspectives to the treatment of nephrolithiasis. Médecine des maladies Métaboliques - Septembre 2013 - Vol. 7 - N°4 doi:10.1016/S1957-2557(13)70593-5

[17] Daudon M, Brahimi M, Nguyenh V, Réveillaud RJ, Jungers P. Intérêt clinique de différencier les phases cristallines de l'oxalate de calcium dans les urines des sujets lithiasiques oxalo-calciques. Acta Pharm Biol Clin 1987 ; 4 : 461-4.

[18] M. Daudon, P. Jungers, B. Lacour Intérêt clinique de l'étude de la cristallurie. Ann Biol Clin 2004, 62 : 379-93

[19] Daudon M. Modèles de cristallisation. In: Jungers P, Daudon M, Le Duc A, editors. Lithiase urinaire. Paris: Flammarion Me ´decine-Sciences; 1989 . p.158–95.

[20] Michel Daudon. Crystalluria. Néphrologie & Thérapeutique 11 (2015) 174–190

[21] Tcheka C, Mbarki M, Loura Benguellah B, Ketcha Mbadcam J, Amechrouq A, Amhoud A, Moubarik A. Apport de l'etude de la cristallurie et du dosage chimique a la recherche des parametres de risque de lacristallisation oxalocalcique urinair E, J Maroc Urol 2012 ; 27-28 : 25-32

[22] Breiman, Leo (2001). "Random Forests". Machine Learning 45 (1): 5–32.

[23] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[24] Cissé Y, Kinouchi Y, Nagashino H, Akutagawa M. Identification of homeostatic dynamics for a circadian signal source using BP neural networks ITBM-RBM 2000; 21: 24 32

[25] Chow TWS, Fei G, Cho SY. High order cumulants-based least squares for nonminimum-phase system identification. IEEE Trans Trans Signal Proc 1997; 45: 104.5-50.

[26] Faraway J, Chatfield C. Time series forecasting with neural network, a comparative study using the airline data. Appl Stat J R Stat Sot Ser 1998; 47: 231-50

[27] Liu Q, Huang T. A neural network with a single recurrent unit for associative memories based on linear optimization. Neurocomputing 118– 2013; 263–267

[28] Kuzmanovski I, Trpkovska M, Šoptrajanov B, Stefov V. Determination of the composition of human urinary stones composed of whewellite, weddellite and carbonate apatite using artificial neural networks Analytica Chimica Acta 491-2003; 211–218

[29] Baxt W G. Use of an artificial neural network for data analysis in clinical decision making: the diagnosis of acute coronary occlusion, neural computation 1990; 2:480-489.

[30] Baxt W G. Analysis of the clinical variables driving decision in an artificial neural network trained to identify the presence of myocardial infarction. Volume 21, Issue 12, December 1992; Pages 1439–1444

[31] Patil S, Henry J W, Rubenfvie M, Stein P D. Neural Network in the clinical diagnosis of acute pulmonary embolism, Chest1993; 104: 1685-1689.

[32] Wu Y, Wu Y, Wang J,Yan Z, Qu L, Xiang B,Zhang Y. An optimal tumor marker group coupled artificial neural network for diagnosis of lung cancer. Expert Systems with Applications 38– 2011; 11329–11334.

[33] Armoni A, Use of neural networks in medical diagnosis, Medical Diagnosis Computing 1998; 15: 100-104

[34] Han M, Snow P B, Brandt J M, Partin A W.Evaluation of artificial neural networks for the prediction of pathologic stage in prostate carcinoma, Cancer 91(S8): 2001; 1661-1666.

[35] HAMILTON D, O'MAHONY D, COFFEY J, MURPHY J, O'HARE N, FREYNE P, WALSH B, COAKLEY D. Classification of mild A lzheimer's disease by artificial neural network analysis of SPET data Nuclear Medicine Communications 1997; 1 8 : 805-810.

[36] Pradhan C,Wuehr M, Akrami F, Neuhaeusser M, Huth S, Brandt T, Jahn K, Schniepp R. Automated classification of neurological disorders of gait using spatio-temporal gait parameters. Journal of Electromyography and Kinesiology 25– 2015; 413–422

[37] Kalman B L, Kwasny S C. Proceeding at the International Joint Conference on Neural Networks, vol. 1, Publisher: IEEE Neural Networks Council, Baltimore,MD, June 1992, pp. 800–805.

[38] Pierna J A F, Lecler B, Conzen J P, Niemoeller A, Baeten V, Dardenne P. Comparison of various chemometric approaches for large near infrared spectroscopic data of feed and feed products: Analytica Chimica Acta 705– 2011; 30– 34.

[39] Baxt W G. Application of artificial neural networks to clinical medicine, The Lancet 346: 1995; 1135-1138

[40] Ait Ider. A, Echajia. M, TCHEKA C, Merbouha A, and Mohamed Mbarki, "Contribution to a diagnosis system of the Whewellite- and Weddellite crystals in human urinary using data minig algorithms," International Journal of Innovation and Applied Studies, vol. 14, no. 2, pp. 297–303, January 2016.

[41] Michel Daudon .La cristallurie : un marqueur diagnostique et pronostique des pathologies cristallogènes et des lithiases rénales. Revue francophone des laboratoires - septembre-octobre 2013 - N°455