

Classification de pathologies du genou en se basant sur la carte de douleur ePKPM

Classification of knee diseases based on the ePKPM pain map

M. Amine Habli¹, N. Mezghani^{2,3}, A. Fuentes², et A. Benazza¹

¹ École Supérieure des Communications de Tunis,
Université de Carthage, laboratoire COSIM,
Cité Technologique des Communications 2083, Ariana, Tunisie.
benazza.Amel@supcom.rnu.tn
mohamed.supcom2012@gmail.com

² Laboratoire de recherche en imagerie et orthopédie
Centre de recherche du CHUM, École de technologie supérieure
Montréal, Québec
neila.mezghani@etsmtl.ca
alexandre.fuentes@gmail.com

³ UER Science et Technologie, TÉLUQ, Montréal, Québec

Résumé Le but de cette étude est de développer un système de classification automatique de pathologies du genou exploitant une version numérisée de la carte photographique de douleur PKPM (ePKPM). La classification s'appuie sur l'algorithme du plus proche voisin pondéré dont les poids sont déterminés selon l'importance des attributs. Les attributs sont des vecteurs descriptifs des régions de douleurs au genou et des symptômes ressentis par le patient. Les performances du système de classification proposé ont été évaluées sur une base de données de 165 patients répartis sur 8 classes de pathologies. Les expériences indiquent que le taux de classification du système est de 74% ce qui démontre la pertinence de la méthode pour évaluer de manière automatique les pathologies considérées en se basant sur les données extraites de la carte ePKPM.

Mots clés Carte électronique de douleur du genou ePKPM, Classification, Plus proches voisins pondérés, pathologies aux genoux.

Abstract The purpose of this study is to develop an automatic classification system of knee pathologies based on the electronic version of the Photographic Knee Pain Map (ePKPM). The classification is performed by the algorithm of weighted nearest neighbor whose weights are adjusted according to the importance of attributes. The set of attributes constitute vectors describing the knee pain regions and the symptoms felt by the patient. The developed classifier considers eight classes of diseases. The performance has been evaluated on a database of 165 patients. The classification rate of the system is 74%, which shows the suitability of the method for automatically evaluating the pathologies considered based on the extracted data from the ePKPM.

Key words Electronic photographic knee pain map ePKPM, Classification, Weighted nearest neighbor, Knee pathologies.

1 Introduction

La douleur est habituellement le symptôme principal des pathologies du genou [3]. Cette sensation est la plus répandue parmi les maladies musculo-squelettiques du membre inférieur touchant une grande partie de la population [9]. En outre elle est signalée par 25% des personnes âgées de plus de 55 ans [7]. L'emplacement et la durée de la douleur au genou ont été déterminés comme étant de bons indicateurs de la structure endommagée [3]. Ces derniers peuvent être évalués avec différents outils tels que les échelles analogiques visuelles qui permettent de juger la gravité de la douleur [4] et les cartes de douleur qui aident à déterminer l'emplacement de la douleur. Plusieurs études ont appréhendé le problème de représentation de la carte de douleurs au genou via un schéma pré-marqué du genou [8], une esquisse schématique divisée en quadrants [1] ou des mannequins du genou [10]. La validité et la fiabilité de ces modèles de cartes ont été souvent démontrées sur des bases de données très réduites ce qui limite la portée de leur utilité en milieu clinique.

Dans une étude récente, un modèle de carte de douleur PKPM (*Photographic Knee Pain Map*) a été proposée par Elson et *al.* pour l'aide au diagnostic, la communication et le développement de projets de recherche [2]. Le PKPM est une photographie d'une paire de genoux subdivisée en zones conçue pour localiser la douleur aux genoux (Figure 1). La validité et la fiabilité en milieu clinique de ce modèle de carte ont été démontrées en utilisant des tests statistiques d'hypothèses [2].

L'objectif de cette étude est d'utiliser une version numérique (ePKPM) de la carte PKPM afin de développer un système automatique de classification de pathologies de genou. Le ePKPM est une interface informatisée qui contient une version numérisée de la carte PKPM à laquelle sont ajoutés des symptômes de pathologies de genou tirés de la littérature [10] et confirmés par les cliniciens collaborant à ce projet de recherche. Le système de classification utilise des données extraites de l'interface ePKPM pour reconnaître de manière automatique la ou les pathologies aux genoux. Ce système est innovateur parce qu'il s'agit de la première étude qui propose un système de classification automatique qui exploite les données décrivant la douleur aux genoux et ses symptômes. Le système proposé devrait offrir, à terme, une meilleure aide au diagnostic de pathologies du genou et une meilleure communication entre le patient et le clinicien.

Le contenu de cet article est organisé comme suit: Dans la section 2, nous décrivons la base de données et nous expliquons l'extraction des caractéristiques à partir de la carte de la douleur ePKPM. Dans la section 3, nous décrivons la

méthode de la classification et nous résumons les résultats expérimentaux obtenus dans la section 4.

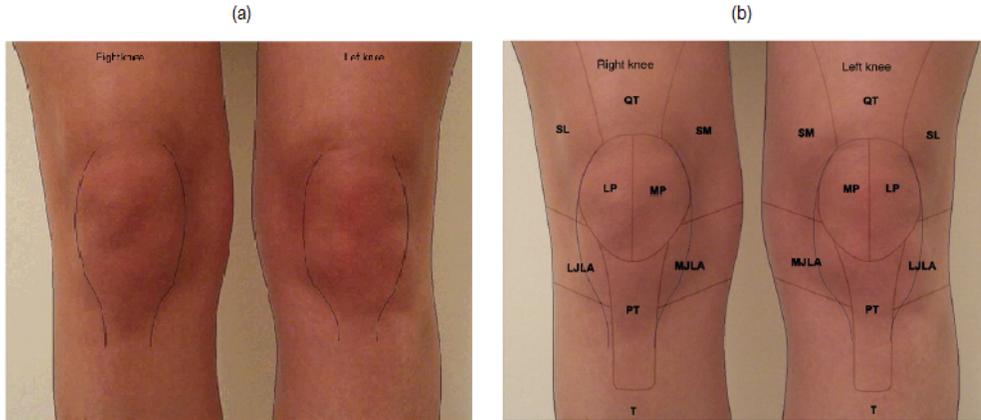


Figure 1. (a) La carte de la douleur PKPM telle que vue par le patient et (b) Les 9 zones superposées sur la carte de douleur: LJLA (ligne latérale de la surface de l'articulation), SL (supérieure latérale), MJLA (ligne médiale de la surface de l'articulation), SM (supérieure médiale), QT (tendon quadriceps), LP (patella latérale), MP (patella médiale), PT (tendon patellaire) et T (tibia). (Images tirées de [2])

2 Extraction des caractéristiques à partir de la carte de la douleur ePKPM

2.1 Base de données

Quatre-vingt-dix-sept sujets ont été recrutés dans le cadre de ce projet. Les sujets ont été examinés au moins une fois par un médecin afin de confirmer leur diagnostic en se basant sur une évaluation clinique, une examination des données radiographiques ainsi qu'une évaluation de la biomécanique des genoux du patient par le système d'analyse de la marche $KneeKG^{TM}$ (Emovi, Canada). Les patients qui ont plusieurs pathologies aux genoux ou qui présentent une douleur diffuse ont été exclus. La base de données ainsi collectée contient 165 genoux répartis sur les 8 classes de pathologies (tel que décrit dans la section 4).

2.2 La carte de la douleur ePKPM

La carte de la douleur PKPM de chaque genou est composée de dix zones: une zone postérieure et de neuf zones antérieures déterminées en se basant sur les struc-

tures anatomiques sous-jacentes du genou et leurs tailles proportionnelles (Figure 1 (b)). Une version électronique (ePKPM) de la carte PKPM a été développée au laboratoire de recherche en imagerie et orthopédie (LIO, de l'École de technologie supérieure, CR-CHUM, Montréal, Canada). Les symptômes cliniques reliés aux pathologies du genou, telles que l'inconfort, l'instabilité et la raideur articulaire, ont été également ajoutés à la version ePKPM (Figure 2).

La carte ePKPM est annotée par le patient en utilisant des marques, en l'occurrence des croix rouges, qui désignent l'emplacement de la douleur sur chaque genou séparément. Le patient sélectionne également l'ensemble des symptômes ressentis à partir d'une liste de symptômes présentée par l'interface.



Figure 2. L'interface d'acquisition de données ePKPM.

2.3 Extraction des caractéristiques

L'extraction des caractéristiques consiste à encoder les informations provenant de la carte ePKPM afin d'assigner à chaque patient l'ensemble des régions marquées, le nombre de marques par région ainsi que l'ensemble des symptômes ressentis. L'extraction des symptômes est une lecture directe des informations recueillies par l'interface d'acquisition des données. Par contre, l'encodage des régions marquées, nécessite la segmentation de l'image de la carte ePKPM de chaque patient selon les 9 régions pour déterminer les régions sélectionnées et le nombre de marques par

régions. Pour ce faire, nous avons converti l'image couleur de la carte de douleur (Fig. 1(b)) en une image à niveau de gris et nous avons procédé à un seuillage. Au final, chaque patient est caractérisé par trois vecteurs de données : un vecteur binaire qui indique la présence/absence de douleur dans chaque région, un vecteur qui indique le nombre de marques dans chaque région et un autre vecteur binaire pour la présence/absence de symptômes.

3 Classification

La classification utilise la méthode du Plus proche voisin pondéré. Cette méthode de classification non paramétrique permet d'assigner à une nouvelle observation la classe d'appartenance de l'échantillon qui lui est le plus proche parmi les échantillons de la base de référence [6]. La similarité est évaluée par des mesures pondérées de distance.

Formellement, soit \mathcal{R} l'ensemble de données de références et soit $X = (x_1, x_2, \dots, x_I)$ de dimension I représentant une forme à classifier. L'algorithme du plus proche se présente ainsi :

- Initialisation des poids w_k selon la relation suivante:

$$\sum_{k=1}^N w_k = 1 \quad (1)$$

N étant le nombre de vecteurs d'attributs (dans notre cas, $N=3$).

- À l'itération n , présentation d'un vecteur de données $X^n = (x_1^n, \dots, x_I^n)$, et calcul de sa distance à chaque échantillons des données de référence $Y^n = (Y_1^n, \dots, Y_I^n)$.

$$(d^n)^2 = \sum_{i=1}^I w_k (x_i - y_i)^2 \quad (2)$$

- Sélection du voisin le plus proche en terme de distance minimale et assignation à l'échantillon X^n la classe de Y^n

Cette méthode peut être très lente en phase de décision. Elle nécessite, en effet, pour un espace d'apprentissage de M échantillons, le calcul de $M - 1$ distances dans un espace à I dimension. Cependant, ce problème ne se pose pas dans cette étude puisque nous disposons d'une petite base de données.

4 Résultats expérimentaux

Le système de classification illustré dans la figure 3 a été validé sur une base de données de 165 genoux en utilisant la validation croisée *Leave one out*. Cette

technique consiste à considérer, à chaque itération, un échantillon de la base de données pour tester et le reste des échantillons comme données de référence. Ce processus est répété sur les 165 genoux de la base pour calculer le taux de bonne classification.

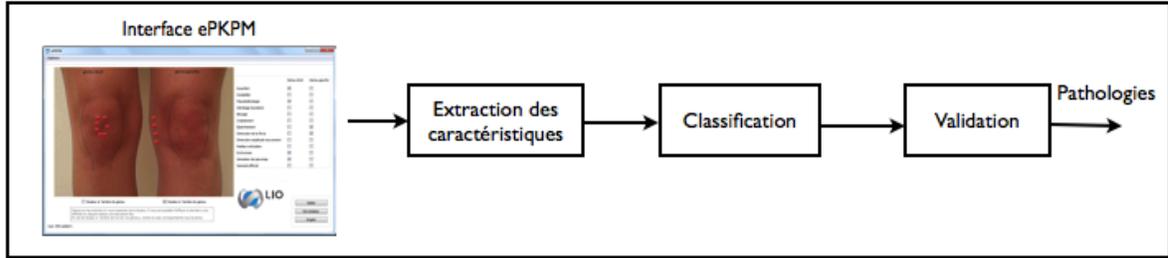


Figure 3. Diagramme du système de classification

Les taux de classification résumés dans la Table 1 sont déterminés en considérant chaque vecteur d'attributs séparément à savoir, la région, la fréquence et les symptômes ainsi que la combinaison pondérée des vecteurs d'attributs. La combinaison des attributs améliore considérablement le taux de classification qui s'élève à 73.9%. Il est noter que la valeur optimale des poids des attributs régions, fréquences et symptômes sont respectivement 0.3, 0.1 et 0.6. Ces valeurs, déterminées empiriquement, démontrent que le système de classification donne plus d'importance à l'attribut symptôme que les autres attributs. Ce résultat corrobore la pratique courante en milieu clinique dans la mesure où les cliniciens les cliniciens accordent beaucoup d'importance aux symptômes pour guider le diagnostic.

Table 1. Taux de classification

Attributs	Taux de classification
Région	55.51%
Fréquence	61.21%
Symptôme	61.21%
Combinaison de la région, des symptômes et des fréquences	73.9%

La visualisation des taux de classifications par classe offre une analyse approfondie des résultats par classe de pathologies (Table 2). Nous remarquons que le taux d'erreur est très important pour certains classes de pathologies. Par exemple, la classe P2 est reconnue à 28.6% des cas. Ceci est dû d'une part au faible nombre

d'échantillons dans cette classe et d'autre part à la grande variabilité des symptômes et de la localisation de la douleur entre les patients atteints d'une même pathologie.

Table 2. Taux de classification par classe de pathologies

Classes	Description de la pathologie	Nombre d'échantillons	$\tau(\%)/$ classe
AS	Asymptomatique	68	100%
P2	Arthrose fémoro-tibiale interne	7	28.6%
P3	Tendinopathie + Syndrome fémuro-rotulien (SFR)	34	70.6%
P4	Lésion du ligament croisé antérieur + lésion du ménisque interne	27	48.1%
P5	Arthrose fémoro-tibiale interne + arthrose fémoro-patellaire	15	33.3%
P6	Arthrose fémoro-patellaire et SFR	4	75%
P7	Arthrose fémoro-tibiale interne + arthrose fémoro-patellaire et SFR	5	80%
P8	Arthrose fémoro-tibiale interne et externe	5	60%

5 Conclusion

Dans cette étude, nous avons développé un système de classification automatique de pathologies du genou en se basant uniquement sur des informations subjectives du patient par l'entremise de l'interface ePKPM. Nous avons testé trois ensemble d'attributs, à savoir la région de douleur, la fréquence et les symptômes ressentis par le patient. Le taux de classification du système développé est de 74% ce qui démontre sa pertinence pour évaluer de manière automatique les pathologies considérées en se basant sur les données extraites de la carte ePKPM.

Cette recherche pourrait être élargie dans les trois directions suivantes. Premièrement, on pourrait considérer une base de données plus large et d'autre pathologies du genou. Deuxièmement, on pourrait envisager une méthode robuste de pondération des attributs. Une telle méthode, par exemple, par algorithmes génétiques [5], permettrait un choix de caractéristiques plus discriminantes, ce qui mènerait donc à un système de classification plus performant. Finalement, on pourrait explorer d'autres attributs comme la fonction biomécanique de l'articulation.

References

1. P. Creamer, M. Lethbridge-Cejku, and M.C. Hochberg. Where does it hurt? pain localization in osteoarthritis of the knee. *Osteoarthritis Cartilage*, 6(5):318–23, 1998.
2. D.W. Elson, S. Jones, N. Caplan, S. Stewart, A.S. Clair Gibson, and D. F D.F. Kader. The photographic knee pain map: Locating knee pain with an instrument developed for diagnostic, communication and research purposes. *The Knee*, 18(6):417–423, 2011.

3. N. Harris and D. Stanley. *Advanced examination techniques in Orthopaedics*. Cambridge University Press, 2003.
4. E.C. Huskisson. Measurement of pain. *Lancet*, 304:1127–31, 1974.
5. A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transaction on Pattern Analysis and Machine intelligence*, 19(2):153–158, 1997.
6. R. Kohavi, P. Langley, and Y. Yun. The utility of feature weighting in nearest-neighbor algorithms. In *Proceedings of the Ninth European Conference on Machine Learning*, pages 85–92. Springer-Verlag, 1997.
7. Croft P. Peat G, McCarney R. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Annals of the Rheumatic Diseases*, 60:91–7, 2001.
8. M. Sengupta, Y. Q. Zhang, J. B. Niu, A. Guermazi, M. Grigorian, D. Gale, D. Felson, and D.J. Hunter. High signal in knee osteophytes is not associated with knee pain. *Osteoarthritis Cartilage*, 14(5):413–7, 2006.
9. M. Urwin, D. Symmons, T. Allison, T. Brammah, H. Busby, M. Roxby, A. Simmons, and G. Williams. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Annals of the Rheumatic Diseases*, 57(11):649–655, 1998.
10. L. R. Wood, G. Peat, E. Thomas, and R. Duncan. Knee osteoarthritis in community-dwelling older adults: are there characteristic patterns of pain location? *Osteoarthritis Cartilage*, 15(6):615–23, 2007.