

# Radiological Pain Predictors in Knee Osteoarthritis, a Four Feature Selection Comparison: Data from the OAI

Jorge I. Galván-Tejada, José M. Celaya-Padilla, Carlos E. Galván-Tejada, Victor Treviño, and José G. Tamez-Peña

Tecnológico de Monterrey, Monterrey, Nuevo León, México  
Bioinformatics  
Escuela de Medicina  
{gatejo,ericgalvan}@uaz.edu.mx,  
{a00811434,vtrevino,jose.tamezpena}@itesm.mx

**Abstract.** In medical science, the image based biomarkers are a recent tool for disease diagnostic and prognostic, the different medical imaging techniques brings a big amount of useful data for analysis and interpretation. The osteoarthritis (OA) is a very common and disabling disease in the industrialized world, pain is the most important and disabling symptom in knee OA, having a preventive treatment is one of the most important tasks in the OA studies. In this work a bioinformatic tool is used to obtain pain prediction models, using genetic algorithms with different feature selection functions multivariate prediction models were obtain and compared based on the medical requirements to investigate radiological features that precede the onset of knee pain, and to identify a radiological-based multivariate prognostic model of knee pain.

## 1 Introduction

Osteoarthritis (OA) is a chronic and complex disease that reduces the quality of life of millions of adults around the world, affecting 1 in every 10 adults over 60 years in the United States [1]. Knee joint pain is the most common and incapacitating symptom [2, 3]. Lifestyle and age of the general population are factors of an increment in the knee OA cases [4-7]. OA remains still poorly understood [8], for a better study and understanding of pain etiology, the Osteoarthritis Initiative (OAI) has been recollecting thousands of clinical data (features) in OA patients; all these features bring important information about the stages of the pain, and other subject characteristics.

Medical imaging is a very important and effective tool to diagnose OA, it is also the most common first hand patient information, using different radiological techniques such as Magnetic resonance imaging (MRI) or X-ray imaging it is possible to obtain a good approach of the OA stage. Differences and advantages of X-ray or MRI techniques have been studied [9-14], it is important to take into account that in emerging countries the first approach to diagnose knee OA is

X-ray imaging, in early stages of OA the use of MRI images in an exploratory stage is not allowed by the monetary cost to the public health systems. This is an important reason to develop a robust prognostic method based on a very extended imaging tool, the X-ray imaging technology. The Osteoarthritis Research Society International (OARSI) assessment scale of knee OA [15, 16], have been adopted as a standard in many health systems and OA studies around the world. Measured by these scales, Joint Space Width (JSW) is commonly associated with OA stages [17, 18], this measurements helps to see the density of the cartilage and the loss of it.

In case of OA, there is no effective treatment other than knee replacement in advanced stages, the pain prediction is important to develop an effective treatment to prevent the disease to reach disabling stages. It is a challenge to correlate the features extracted from x-ray imaging and OA symptoms as has been presented in other works [19-22]. An exploratory computational analysis of medical data can contribute to get a better understanding in some diseases, prevent some disorders associated with early stages, and in some risk subjects can contribute to the improvement of health systems around the world, computational statistics and Bioinformatics tools can be a powerful way to find some unexpected connections between some subjects characteristics and the symptoms of a certain disease, the objective of bioinformatics was to explore the human genome, but have shown to be effective in the association of subjects characteristics and symptoms in other diseases [24-27].

Using GALGO[23] an R package based on genetic algorithms bioinformatics tool, five different classification methods; logistic regression; linear discriminant; nearest centroid; KNN, were used to obtain a pain prediction model, with all the clinical and radiological information available in the OAI databases, a yearly observed case/control group of subjects was used to predict the pain development based on X-ray assessed features.

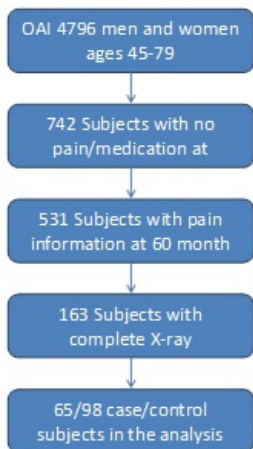
The objective of this work is to determine if using different classification methods in bioinformatics tools we can have a better correlation between the features obtained from the knee radiological features and the future pain symptom.

## 2 Methods

### 2.1 Study Design

In this work, a case/control group was defined from OAI databases was selected based on the right knee radiological assessments, only subjects with no radiological or clinical missing data were included in this study, 163 (93) subjects (females). The criteria inclusion for all subjects was; not presenting pain as a symptom; not presenting a symptomatic status; and taking no pain medication at the baseline visit; all control and case should accomplish the inclusion criteria, the selection process is presented in Figure 1. Of all these subjects, having pain as a chronic symptom at some timepoint after their baseline and up to their 60 month visit were considered cases. The rest, the ones with no pain and no pain medicated from their baseline until their 60 month visit, were considered

controls. The demographic information is shown in Table 1. Looking for a pain prediction model, the one year before pain incidence radiological assessments were used as predictive variables (T-1).



**Fig. 1.** Subject Selection Process

**Table 1.** Demographic characteristics of study participants

SD	Case	Control	All
Subjects (Females)	65 (38)	98 (55)	163 (93)
Average height (SD) [m]	1.69 (.09)	1.68 (.1)	1.68 (.09)
Average BMI (SD)	27.05(4.3)	26.27 (4)	26.58 (4.1)
Average age (SD)	62.69 (9.6)	61.80 (10.1)	62.15 (9.9)
Age range	46 - 78	45 - 79	45 - 79

Datasets used in this work were: Central Assessment of Longitudinal Knee X-rays for Quantitative JSW ver. 1.6, from which quantitative data were obtained, and Right knee symptom status, from which the predicted variable, chronic pain, was obtained. This information was pre analyzed by a radiologist group associated to the OAI using the OARSI quantitative grading scale [28, 29][30-32].

## 2.2 Statistical Analysis

For statistical analysis, seventeen quantitative features were measured in right knee radiographs, the description of the features is shown in table 2. Since men are generally taller, all image features from the quantitative datasets went through a height and gender adjustment using a linear regression [33]:

$$JSW_{adj} = JSW - b_0 - (Height * b_1) - (Gender * b_2) \quad (1)$$

Where  $JSW_{adj}$  represents the adjusted measurement,  $JSW$  is the original measurement, and  $b_0$ ,  $b_1$  and  $b_2$  are the coefficients obtained from the linear regression. The values assigned to the Gender variable were 0 for males and 1 for females. The data collected was Z normalized using the rank-based inverse normal transform.

**Table 2.** Features and description

Feature ID	Description
MCMJSW	Medial minimum JSW [mm]
JSW175	Medial JSW at x=0.175 [mm]
JSW200	Medial JSW at x=0.200 [mm]
JSW250	Medial JSW at x=0.250 [mm]
JSW300	Medial JSW at x=0.300 [mm]
JSW225	Medial JSW at x=0.225 [mm]
JSW150	Medial JSW at x=0.150 [mm]
JSW275	Medial JSW at x=0.275 [mm]
LJSW850	Lateral JSW at x=0.850 [mm]
LJSW900	Lateral JSW at x=0.900 [mm]
LJSW700	Lateral JSW at x=0.700 [mm]
LJSW825	Lateral JSW at x=0.825 [mm]
LJSW750	Lateral JSW at x=0.750 [mm]
LJSW875	Lateral JSW at x=0.875 [mm]
LJSW725	Lateral JSW at x=0.725 [mm]
LJSW800	Lateral JSW at x=0.800 [mm]
LJSW775	Lateral JSW at x=0.775 [mm]

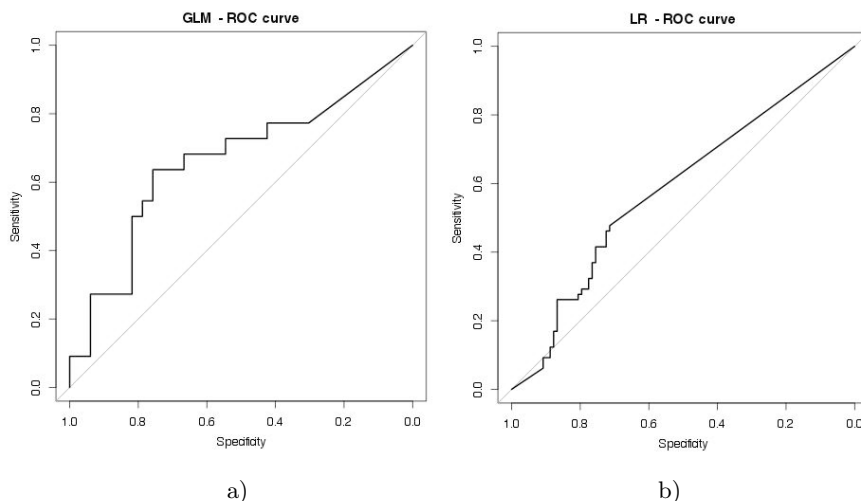
After data transformation and normalization, four multivariate searches were then performed, using the T-1. For all searches, the 60 month visit right knee pain information was used as the outcome variable. These searches were performed using Galgo. The algorithm was set to return 1000 predictive models, with 5 features each. Each one of these models evolved throughout 600 generations, replicating, recombining and mutating from a set of models with random features, optimizing for its accuracy, obtained using a 3-k fold cross validation with a 2/3-1/3 train-test proportion. Using the same seed to train and test, four classification strategies were used (Logistic regression, linear discriminant, KNN and nearest centroid), one on each model, for all searches the best prediction model was refined using the same strategy: Features were then ranked according to their frequency in the 1000 models, and using such rank, a Forward Selection strategy was carried out, prompting a new predictive model. Finally, the size of this model was shrunk by using a Robust Gene Backward Feature Elimination algorithm, in which features were removed if their contribution to the accuracy of the model, measured using the same cross-validation strategy. For all four predictive models the Area Under Receiver Operating Characteristics (AUROC) was calculated to obtain the specificity vs sensitivity value. A

univariate correlation with the outcome of the data was performed to test the correlation of each feature with the outcome of pain prediction. All statistical analysis was done using R software [34].

### 3 Results

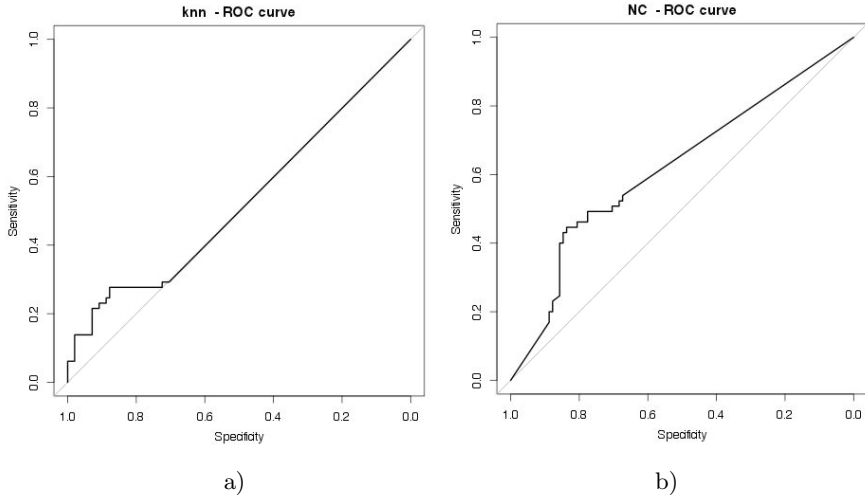
The obtained results:

- For logistic regression as cost function the best performing model was a four feature model, obtained an AUCROC of 0.6612, the curve is shown in figure 2a.
- For Linear Discriminant, the best performing predictive model was a two feature model, obtained an AUCROC of 0.5845, the curve is shown in figure 2b.
- For Knn, the best performing model was a two feature model, obtained an AUCROC of 0.5232, the curve is presented in figure 3a.
- For Nearest centroid experiment, the best performing model was a five feature model, obtained an AUCROC of 0.6216, the curve is presented in figure 3b.



**Fig. 2.** ROC Curves a) Logistic Regression b) Linear Discriminant

In table 3, the description of the predictive models and its features is presented. In the univariate experiment, as expected the correlation was minimal, due the high variability of the data, nevertheless, the top 10 rank correlated features is presented in table 4 as the absolute correlated value of the feature.



**Fig. 3.** ROC Curves a)Knn b)Nearest Centroid

**Table 3.** Predictive models description

Classifier	Logistic regression	linear discriminant	Knn	Nearest Centroid
Features	AV06JSW300, AV06JSW150, AV06LJSW750	AV06JSW275 AV06LJSW750	V06JSW300 V06LJSW775	AV06JSW300 AV06LJSW750 AV06LJSW875 AV06LJSW775 AV06LJSW800

**Table 4.** Top ten features correlated to the Outcome

Feature	Correlation to Outcome
AV06LJSW875	0.2088991
AV06JSW275	0.2067434
AV06JSW300	0.1998420
AV06LJSW800	0.1829875
AV06LJSW825	0.1827017
AV06LJSW900	0.1820316
AV06LJSW850	0.1819792
AV06LJSW775	0.1666364
AV06JSW250	0.1645778
AV06JSW200	0.1531015

## 4 Conclusions

Based on the experiment results, the logistic regression and nearest centroid models showed a better performance in sensitivity and specificity, is important

to say, these models are a tool to find symptoms as risk factors to develop knee pain; this way, the features in the predictive models are features associated with advanced stages in the disease, if some early changes in these measures can be predictive a year before of the pain appearance it could be helpful for preventing treatment development.

Due to its nature, pain is complicated to measure; nevertheless, the data appears to have a good performance in pain prediction. Some early radiological features can be associated with OA symptoms; it is possible to achieve future pain prediction with multivariate models based on X-ray features. The association of future pain with X-ray features can be used to develop a prevention strategy for chronic pain. The public health systems can be beneficial with the use of a radiological Biomarker.

In other hand, the computational analysis of data sets can be very helpful to develop image based biomarkers, due to its nature, this analysis can manage a big datasets and perform millions of operations to correlate some subject factors with disease symptoms and this way have a better and faster diagnosis.

The present study shows a close relation between the X-ray based models obtained in the multivariate regression and the development of clinical pain associated with the disease, which strongly suggest the feasibility of a Bio-Marker based in those features, this Bio-Marker can be used to aid the radiologist in the process of diagnose the patient or a second opinion, this represent a great area of impact specially in developing countries, in which access to the high level health care system is very restricted.

**Acknowledgements.** This work was partially supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT), by Grant 16864 Ciencia Bsica from CONACYT, and by Catdra de Bioinformtica (CAT220) from Tecnológico de Monterrey. J. I. G.T. and C. E. G. T. thanks to PROMEP for partially support his doctoral studies. “The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.”

## References

1. White, D.K., TudorLocke, C., Felson, D.T., Gross, K.D., Niu, J., Nevitt, M., Lewis, C.E., Torner, J., Neogi, T.: Do radiographic disease and pain account for why people with or at high risk of knee osteoarthritis do not meet physical activity guidelines? *Arthritis & Rheumatism* 65, 139–147 (2013)

2. Neogi, T.: The epidemiology and impact of pain in osteoarthritis. *Osteoarthritis Cartilage* 21, 1145–1153 (2013)
3. Agaliotis, M., Fransen, M., Bridgett, L., Nairn, L., Votrubec, M., Jan, S., Heard, R., Mackey, M.: Risk factors associated with reduced work productivity among people with chronic knee pain. *Osteoarthritis Cartilage* 21, 1160–1169 (2013)
4. Neogi, T., Bowes, M.A., Niu, J., De Souza, K.M., Vincent, G.R., Goggins, J., Zhang, Y., Felson, D.T.: Magnetic resonance imaging-based three-dimensional bone shape of the knee predicts onset of knee osteoarthritis: data from the osteoarthritis initiative. *Arthritis Rheum.* 65, 2048–2058 (2013)
5. Colbert, C.J., Almagor, O., Chmiel, J.S., Song, J., Dunlop, D., Hayes, K.W., Sharma, L.: Excess body weight and four-year function outcomes: comparison of African Americans and whites in a prospective study of osteoarthritis. *Arthritis Care Res. (Hoboken)* 65, 5–14 (2013)
6. Riddle, D.L., Stratford, P.W.: Body weight changes and corresponding changes in pain and function in persons with symptomatic knee osteoarthritis: a cohort study. *Arthritis Care Res. (Hoboken)* 65, 15–22 (2013)
7. Tanamas, S.K., Wluka, A.E., Davies-Tuck, M., Wang, Y., Strauss, B.J., Proietto, J., Dixon, J.B., Jones, G., Forbes, A., Cicuttini, F.M.: Association of weight gain with incident knee pain, stiffness, and functional difficulties: a longitudinal study. *Arthritis Care Res. (Hoboken)* 65, 34–43 (2013)
8. Abhishek, A., Doherty, M.: Mechanisms of the placebo response in pain in osteoarthritis. *Osteoarthritis Cartilage* 21, 1229–1235 (2013)
9. Guermazi, A., Niu, J., Hayashi, D., Roemer, F.W., Englund, M., Neogi, T., Aliabadi, P., McLennan, C.E., Felson, D.T.: Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). *BMJ* 345, e5339 (2012)
10. Wirth, W., Duryea, J., Hellio Le Graverand, M.-P., John, M.R., Nevitt, M., Buck, R., Eckstein, F.: Direct Comparison of Fixed Flexion Radiography and MRI in Knee Osteoarthritis: Responsiveness Data from the Osteoarthritis Initiative. *Osteoarthritis and Cartilage* (2012)
11. Cotofana, S., Wyman, B.T., Benichou, O., Dreher, D., Nevitt, M., Gardiner, J., Wirth, W., Hitzl, W., Kwok, C.K., Eckstein, F., Frobell, R.B., Group, O.I.: Relationship between knee pain and the presence, location, size and phenotype of femorotibial denuded areas of subchondral bone as visualized by MRI. *Osteoarthritis Cartilage* 21, 1214–1222 (2013)
12. Baert, I.A., Staes, F., Truijzen, S., Mahmoudian, A., Noppe, N., Vanderschueren, G., Luyten, F.P., Verschueren, S.M.: Weak associations between structural changes on MRI and symptoms, function and muscle strength in relation to knee osteoarthritis. *Knee Surg Sports Traumatol Arthrosc* (2013)
13. Neogi, T., Felson, D., Niu, J., Nevitt, M., Lewis, C.E., Aliabadi, P., Sack, B., Torner, J., Bradley, L., Zhang, Y.: Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. *BMJ* 339, b2844 (2009)
14. Haugen, I.K., Slatkowsky-Christensen, B., Boyesen, P., van der Heijde, D., Kvien, T.K.: Cross-sectional and longitudinal associations between radiographic features and measures of pain and physical function in hand osteoarthritis. *Osteoarthritis Cartilage* 21, 1191–1198 (2013)
15. Kellgren, J., Lawrence, J.: Radiological assessment of osteo-arthritis. *Annals of the Rheumatic Diseases* 16, 494 (1957)
16. Altman, R.D., Gold, G.: Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis and Cartilage* 56, A1–A56 (2007) (revised)



17. Bloecker, K., Guermazi, A., Wirth, W., Benichou, O., Kwok, C.K., Hunter, D.J., Englund, M., Resch, H., Eckstein, F., Investigators, O.: Tibial coverage, meniscus position, size and damage in knees discordant for joint space narrowing - data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* 21, 419–427 (2013)
18. Attur, M., Krasnokutsky-Samuels, S., Samuels, J., Abramson, S.B.: Prognostic biomarkers in osteoarthritis. *Curr. Opin. Rheumatol.* 25, 136–144 (2013)
19. Kinds, M.B., Marijnissen, A.C., Bijlsma, J.W., Boers, M., Lafeber, F.P., Welsing, P.M.: Quantitative radiographic features of early knee osteoarthritis: development over 5 years and relationship with symptoms in the CHECK cohort. *J. Rheumatol.* 40, 58–65 (2013)
20. Glass, N.A., Torner, J.C., Frey Law, L.A., Wang, K., Yang, T., Nevitt, M.C., Felson, D.T., Lewis, C.E., Segal, N.A.: The relationship between quadriceps muscle weakness and worsening of knee pain in the MOST cohort: a 5-year longitudinal study. *Osteoarthritis Cartilage* 21, 1154–1159 (2013)
21. Shimura, Y., Kurosawa, H., Sugawara, Y., Tsuchiya, M., Sawa, M., Kaneko, H., Futami, I., Liu, L., Sadatsuki, R., Hada, S., Iwase, Y., Kaneko, K., Ishijima, M.: The factors associated with pain severity in patients with knee osteoarthritis vary according to the radiographic disease severity: a cross-sectional study. *Osteoarthritis Cartilage* 21, 1179–1184 (2013)
22. Hochman, J.R., Davis, A.M., Elkayam, J., Gagliese, L., Hawker, G.A.: Neuropathic pain symptoms on the modified painDETECT correlate with signs of central sensitization in knee osteoarthritis. *Osteoarthritis Cartilage* 21, 1236–1242 (2013)
23. Trevino, V., Falciani, F.: GALGO: an R package for multivariate variable selection using genetic algorithms. *Bioinformatics* 22, 1154–1156 (2006)
24. Galván-Tejada, J., Martínez-Torteya, A., Totterman, S., Farber, J., Treviño, V., Tamez-Pena, J.: A wide association study of predictors of future knee pain: data from the osteoarthritis initiative. *Osteoarthritis and Cartilage* 20, S85 (2012)
25. Martínez-Torteya, A., Galván-Tejada, J., Totterman, S., Farber, J., Treviño, V., Tamez-Pena, J.: Can T2 relaxation be used to predict koos other symptoms?-data from the osteoarthritis initiative. *Osteoarthritis and Cartilage* 20, S208–S209 (2012)
26. Martínez-Torteya, A., Treviño-Alvarado, V.M., Tamez-Peña, J.G.: Improved multimodal biomarkers for Alzheimer’s disease and mild cognitive impairment diagnosis: data from ADNI. In: *SPIE Medical Imaging*, pp. 86700S–86700S–86709. International Society for Optics and Photonics (Year)
27. Torteya, A.M., Tamez Peña, J.G., Treviño Alvarado, V.M.: Multivariate predictors of clinically relevant cognitive decay: A wide association study using available data from ADNI. *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association* 8, P285–P286 (2012)
28. Neumann, G., Hunter, D., Nevitt, M., Chibnik, L., Kwok, K., Chen, H., Harris, T., Satterfield, S., Duryea, J.: Location specific radiographic joint space width for osteoarthritis progression. *Osteoarthritis and Cartilage* 17, 761–765 (2009)
29. Duryea, J., Li, J., Peterfy, C., Gordon, C., Genant, H.: Trainable rule-based algorithm for the measurement of joint space width in digital radiographic images of the knee. *Medical Physics* 27, 580 (2000)
30. Kellgren, J., Lawrence, J.: *Atlas of standard radiographs: the epidemiology of chronic rheumatism, vol. 2.* Blackwell, Oxford (1963)
31. Felson, D.T., Nevitt, M.C., Yang, M., Clancy, M., Niu, J., Torner, J.C., Lewis, C.E., Aliabadi, P., Sack, B., McCulloch, C.: A new approach yields high rates of radiographic progression in knee osteoarthritis. *The Journal of Rheumatology* 35, 2047–2054 (2008)

32. Hing, C.B., Harris, M.A., Ejindu, V., Sofat, N.: The Application of Imaging in Osteoarthritis
33. Suri, P., Hunter, D.J., Rainville, J., Guermazi, A., Katz, J.N.: Presence and extent of severe facet joint osteoarthritis are associated with back pain in older adults. *Osteoarthritis Cartilage* 21, 1199–1206 (2013)
34. <http://www.r-project.org/>