# Application of Machine Learning Decision Tree in Diagnosing Joint Pain

Saman SHAHID [1],
[1] Department of Sciences & Humanities, National University of Computer & Emerging Sciences (NUCES), Lahore Pakistan.
Email: drshahidsaman@gmail.com

Aatir JAVAID [2]
[2] Department of Orthopedic Surgery, University College of Medicine and Dentistry (UCMD), Lahore Pakistan.

***Abstract-*** Pain is widely acknowledged to be a complicated experience. Pain is also a symptom of joint inflammation in arthritis, such as rheumatoid arthritis and osteoarthritis. Machine learning (ML) for classification uses nonparametric supervised learning techniques called decision trees. By gaining knowledge from decision rules created from the characteristics of the supplied data, they may be utilized to forecast the target variable. We applied a machine learning decision model to anticipate joint pain (target variable) while considering clinical biochemistry. A total of 650 patients were included who visited orthopedic OPD with joint swelling or myalgia. Decision Tree was trained tested, and cross-validated with supervised learning. The model was evaluated along with the selected features/attributes (age, gender, uric acid, & CRP). 44% of patients were diagnosed with joint pain. The decision tree model yielded an accuracy of 94% and a validation accuracy of 96%. Uric acid was strongly correlated with joint pain. Early ML-based joint pain identification will help avert more significant orthopedic issues.
***Keywords:*** Joint Pain; Arthritis; Machine Learning; Decision Tree; C-Reactive Protein; Uric Acid

## I. INTRODUCTION

Injury to any of the ligaments, bursae, or tendons around the joint might result in joint pain. Injury to the ligaments, cartilage, and bones within the joint can also occur. Pain is also a symptom of joint inflammation (arthritis, rheumatoid arthritis, and osteoarthritis) and infection, and it can be a cause of joint cancer in rare cases. Joint pain is accompanied by the following indications and symptoms: joint redness, swelling, soreness, warmth, limping, freezing of the joint, stiffness, weakness, and so on [1].

Complex clinical and experimental data may be used by modern computer science methods to better understand the intricacies of pain. In the United States, 24% of individuals suffer from arthritis. It is the main factor causing job disabilities [3]. Knee osteoarthritis is the most common among these people, affecting 95% of them, either as a single joint or in combination with other joints such as the hip joint [2].

A kind of arthritis known as gout is characterized by elevated uric acid levels in the blood. In this condition, uric acid crystals form and assemble in the synovial fluids. The intricate mechanisms that regulate the hepatic production of uric acid as well as its renal and gastrointestinal excretion are controlled by several factors. Uric acid is produced as a result of the metabolism of both endogenous purines and an external purine pool [4].

The body's inflammation can be detected with the C-reactive protein (CRP) test. There is a strong link between CRP and osteoarthritis of the knee as a measure of acute phase response and inflammation and the patients may feel discomfort in their joints, musculoskeletal pain, arthritis, or arthralgia [5].

The goal of the current study was to determine if a machine learning decision model could accurately forecast joint pain in conditions of clinical biochemistry. Early diagnosis of joint pain can aid in preventing more severe orthopedic issues such as osteoarthritis, fibromyalgia, reactive arthritis, septic arthritis, gout, osteonecrosis, bursitis, and osteomyelitis, among others.

We included patients for this study who attended the orthopedic department of Mansoorah Hospital Lahore, Pakistan and had minor joint edema or myalgia. Using clinical symptoms or risk indicators as inputs, machine learning (ML) via the decision-based algorithm is a valuable technique for making clinical predictions. Therefore, the ML-based decision/prediction models can help doctors with patients' optimal quality of life for refined diagnosis and selection of suitable therapy for the relevant clinical illness.

Machine learning (ML) for classification uses nonparametric supervised learning techniques called decision trees. By gaining knowledge from decision rules created from the characteristics of the supplied data, they may be utilized to forecast the target variable. Alexos et al. (2020) [6] applied ML models and presented an approach that exhibited exceptional potential for detecting pain development at an early stage, hence boosting future knee osteoarthritis preventive efforts.

Uric acid has been shown by Kono et al. (2010) [8] to be a physiological regulator of the inflammation brought on by tissue damage. Uric acid in synovial fluid influences osteoarthritis tissue inflammation, disease severity, and progression [9]. About 40% of persons over the age of 70 have osteoarthritis, a persistent, painful joint condition [10]. Alexos et al. (2020) [6] used baseline data to create a tool that predicted the evolution of pain in knee osteoarthritis patients. They applied ML models on various combinations of feature subsets, yielding results of up to 84.3 percent with only a modest number of features. Chan et al. (2021) [11] evaluated the contributions of each local and systemic risk factor in the multi-etiology of knee osteoarthritis to disease start and worsening using a unique mix of machine learning methodologies. Except for medial joint space narrowing, history of injury has the highest DeepLIFT gradient for knee osteoarthritis onset prediction. Kokkotis et al. (2020) [12] examined multidisciplinary data from the osteoarthritis initiative database. The validated criteria were validated in data subgroups using seven well-known classifiers in five distinct techniques. Support Vector Machine obtained a classification accuracy of 74 percent on the first 55 risk

978-1-6654-5432-2/22/$31.00 ©2022 IEEE

variables tested. The findings served as the foundation for the creation of accurate techniques for predicting knee osteoarthritis progression.

## II. MATERIALS AND METHODS

### A. *Study Design & Selection Criteria:*

Patients included with aged between 18 to 70 years visited the orthopedic clinic with mild joint swelling or myalgia. We excluded patients with co-morbidities (hypertension, diabetes and heart diseases, etc.), degenerative joint changes (osteoarthritis), or autoimmune diseases (rheumatoid arthritis, SLE- systemic lupus erythematosus). The patients were examined by the orthopedician to include/exclude patients.

### B. *Data Collection & Statistical Analysis:*

Biochemistry laboratory tests were done in the biochemistry lab at Mansoorah Hospital, Lahore. For CRP measurements, the Turbilatex test kit was used and for uric acid, the Benecheck uric acid test kit was used. A Pearson bi-serial correlation analysis was conducted to analyze the significant correlations between the clinical variables and the presence/absence of joint pain.

### C. *Machine Learning Decision Tree:*

The decision tree model was constructed using the attributed dataset of patients treated outside. A predictive decision tree model uses observations about a parameter to create branches that conclude the parameter's targets. The remaining instances might be categorized by going from the root of the tree to its leaf nodes. While every branch of the tree falls, every node displays attribute testing of an instance. The model was evaluated along with the selected features. Age, gender, uric acid, and CRP were included as independent factors, and the dependent variable diagnostic for the presence or absence of joint pain was used to determine an individual's result. The decision tree rule was used to create a classification decision tree model. In both classification and regression, this rule divided values into different groups to reduce error as effectively as possible given the parameter requirement. While an estimation for a numerical value was produced by averaging the values in a leaf, a prediction for the class label attribute was discovered by doing so [7]. Table 1 shows the characteristics and parameters considered for all decision-based trees. The dataset was split into 70% training and 30% testing datasets. The attribute 'diagnosis' (joint-pain) was the target, labeled as 'yes/no'.

### D. *Generation of Decision Models:*

The model was constructed using the subsequent steps: The process involved retrieving/accessing the dataset, setting the target attribute (diagnosis), assigning training/testing portions of the dataset as inputs for the model (decision tree), training the model with supervised learning algorithms, creating the model, applying it to the testing dataset, and evaluating the statistical performance of the models. The following are included: correlations, accuracies, kappa, weighted mean precision, absolute error, relative error, root mean squared error, and others. To validate the output of the trained models, 10-fold cross-validation was also investigated.

## III. RESULTS

### A. *Background Clinical Correlation:*

There were 53.5% male and 46.5% female patients. The mean age of the patients was 52 years. 44 patients were diagnosed with joint pain, whereas. Only uric acid was found positively strong correlated (p-value: 0.000) with the occurrence of joint pain

### B. *Classification Decision-Trees & Validation:*

Table 2 enlists the details of all ML trees for accuracies, correlations, class recalls, root relative squared errors, root mean squared errors, absolute errors, and relative errors both for models as well as for 10-fold cross-validations. The decision tree model yielded an accuracy of 94.36% and a validation accuracy of 96.46%.

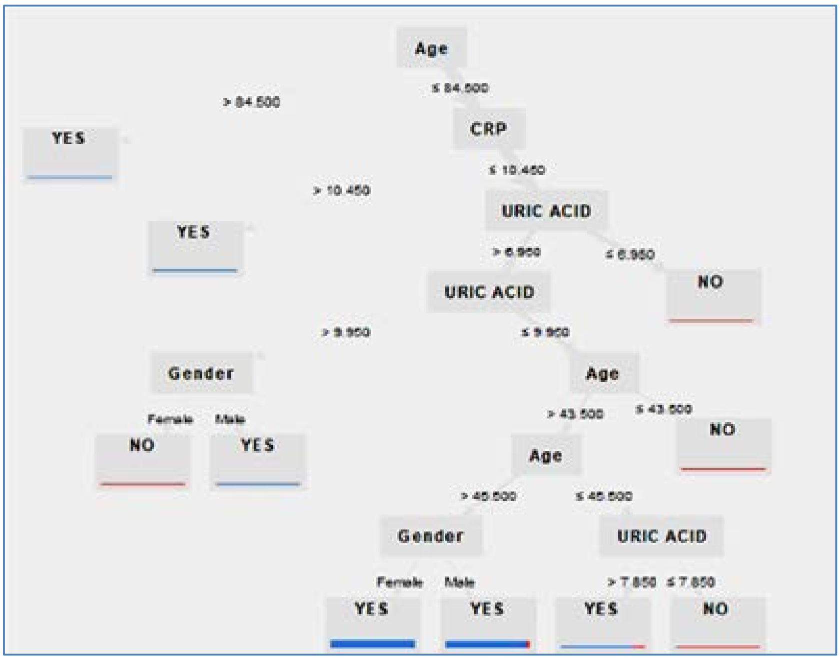


FIGURE 1. Decision Tree Forest

978-1-6654-5432-2/22/$31.00 ©2022 IEEE

TABLE 1. DECISION TREE CHARACTERISTICS

| **Model** | Training Ratio | 0.7 | Testing Ratio | 0.3 |
|---|---|---|---|---|
| **Decision Tree** (Pruning applied); Attribute: Diagnosis; Target Role: Label (outcome); automatic sampling | Criterion: Gain Ratio | 72% | Minimal Gain | 0.01 |
| | Max. Depth | 10 | Minimal leaf size | 2 |
| | Confidence | 0.1 | Minimal size for split | 4 |

TABLE 2. MACHINE LEARNING – DECISION MODEL

| **Model Accuracies and Correlations** | Errors for Model [a] and Cross Validations [b] | | | | |
|---|---|---|---|---|---|
| **Decision Tree** Accuracy: 94%; Correlation: 0.938 Kappa: 0.885 10-Fold Cross-validation Accuracy: 96%; Correlation: 0.938 Kappa: 0.982 | **Root Relative Squared Error** | **Root Mean Squared Error** | **Absolute Error** | **Relative Error** | **Class Precision** |
| | [a]0.389 | [a] 0.172 +/-0.000 | [a] 0.059 +/-0.179 | [a] 5.92 +/-17.92% | 97.47% (true yes) |
| | [b]0.327+/-0.227 | [b]0.145+/-0.101[b] | [b]0.047+/-0.028 | [b]4.67+/-2.77% | 92.24% (true no) |

## IV. DISCUSSION

The study analyzed the machine learning decision model that can predict joint pain in clinical variables. The presence of joint discomfort was shown to be strongly positively linked with uric acid. The dependent variable diagnostic for the presence or absence of joint pain was utilized to predict an individual's outcome. Age, gender, uric acid, and CRP were added as independent variables. A classification decision tree model was made using the decision tree rule. This rule separated data into many categories in both classification and regression to decrease error as effectively as feasible given the parameter need.

The decision tree model performed well with an accuracy of 94%. As seen in Fig. 1, the decision tree first analyzes the patient's age and creates classes for further evaluation of risks for joint pain (yes/no) such as levels of CRP and uric acid. It also assesses the uric acid levels concerning gender (male/female) as a risk factor of having joint pain prediction.

Aljaaf et al. (2016) [13] estimated the frontal plane internal knee abduction moment from 3D Euler angles of the ankle, knee, hip, and pelvis during a single gait cycle of 31 individuals with alkaptonuria to assess the feasibility of this technique. In terms of prediction performance, the random forest technique performed the best, but it was also the slowest by a factor of 10. Age-related progression of temporomandibular joint osteoarthritis' chronic damage makes it crucial to identify the condition early on before morphological deterioration happens. Bianchi et al. (2020) [14] included four machine learning models in the test: Logistic Regression, Random Forest, LightGBM, and XGBoost. The XGBoost + LightGBM models had an accuracy of 0.823 for diagnosing temporomandibular joint osteoarthritis.

Early diagnosis of individuals with chronic illnesses who are at risk of developing persistent pain is therapeutically desired to begin multimodal therapy on time. In a study, machine learning was utilized to find early factors that predict the development of chronic pain in rheumatoid arthritis. Swollen joint count and painful joint count obtained at 3 months showed a balanced accuracy of rheumatoid arthritis of 59 percent when predictors were restricted to objective clinical markers of disease severity. The results showed that machine learning is well adapted to extracting knowledge from data retrieved from pain and illness registries. Early functional characteristics of rheumatoid arthritis are predictive of the severity and duration of chronic pain [15]. Yoo et al. (2017) [16] did research to predict rheumatoid arthritis patients from the k-Means algorithm. Clarity evaluation scores of 84 percent or above were obtained using the explanatory model. Rheumatoid arthritis significantly lowers the quality of life for patients in an aging population. However, by using the results of this study to predict rheumatoid arthritis, we may be able to improve people's quality of life.

Ichikawa et al. (2016) [17] developed a machine learning-based identification system to diagnose the high risk of hyperuricemia. The used gradient boosted random forest trees and checked their accuracies to develop a virtual health check-up system in Japan.

## V. CONCLUSION

Pain is typically recognized as a complex feeling. Pain is also a sign of arthritic joint inflammation, such as rheumatoid arthritis and osteoarthritis. We used a machine learning decision model to predict joint pain (the goal variable) while taking clinical biochemistry into account. A total of 650 individuals with joint swelling or myalgia who visited an orthopedic OPD were included in the study. With supervised learning, Decision Tree was trained, tested, and cross-validated. The model was assessed in conjunction with the chosen features/attributes (age, gender, uric acid, & CRP). Uric acid was shown to be strongly associated with the prevalence of joint discomfort. As independent factors, age, gender, uric acid, and CRP were included. The decision tree rule was used to create a classification decision tree model. With an accuracy of 94%, the decision tree model worked admirably.

Because of the diverse range of potential causes and treatment effects, patients with back, knee, or ankle pain are common and present challenges in routine medical practice. We can demonstrate that identifying the underlying diagnostic categories requires a mix of different AI methods and baseline data. Early AI-based diagnosis of joint discomfort will help avert more serious orthopedic issues. We anticipate advancing research into patient-specific treatment strategies for osteoarthritis and enhancing the condition of articular joints as a result. To supplement physicians' decision-making and to increase the present data collection, additional patient data are urgently required. It should be noted that these models can be accessed by physicians via. a deployed model via online services or web-based apps. These models' platforms can be applied in hospitals using integrated computerized medical systems.

The authors report no conflict of interests.

978-1-6654-5432-2/22/$31.00 ©2022 IEEE

ACKNOWLEDGMENTS:
We are thankful to the biochemistry laboratory of Mansoorah Hospital for the provision of the data.

REFERENCES:

[1] W.C. Shiel. MedicineNet. “Joint Pain: Signs & Symptoms”. 2019. Available from: https://www.medicinenet.com/joint_pain/symptoms.htm
[2] E. Akhter, S. Bilal, U. Haque. “Prevalence of arthritis in India and Pakistan: a review”. Rheumatology international. 2011, vol. 31(7), pp. 849-55.
[3] National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). Centers for Disease Control & Prevention (CDC). Arthritis. 2021. Available from:
[7] J. Lötsch, L. Alfredsson, J. Lampa. “Machine-learning–based knowledge discovery in rheumatoid arthritis–related registry data to identify predictors of persistent pain”. Pain. 2020, vol. 161(1), pp. 114-26.
[8] J. Yoo, M.K. Li, C. Ihm, E.S. Choi, M.S. Kang. “A study on prediction of rheumatoid arthritis using machine learning”. International Journal of Applied Engineering Research. 2017, vol. 12(20), pp. 9858-62.
[9] D. Ichikawa, T. Saito, W. Ujita, H. Oyama. “How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach.” Journal of biomedical informatics. 2016, vol. 64, pp. 20-4.
[10] RapidMiner9. Available from: https://docs.rapidminer.com/latest/studio/operators/rapidminer-studio-operator-reference.pdf
[11] H. Kono, C.J. Chen, F. Ontiveros, K.L. Rock. “Uric acid promotes an acute inflammatory response to sterile cell death in mice”. The Journal of clinical investigation. 2010, vol. 120(6), pp. 1939-49.
[12] A.E. Denoble, K.M. Huffman, T.V. Stabler, S.J. Kelly, M.S. Hershfield, G.E. McDaniel, R.E. Coleman, V.B. Kraus. “Uric acid is a danger signal of increasing risk for osteoarthritis through inflammasome activation”. Proceedings of the National Academy of Sciences. 2011, vol. 108(5), pp. 2088-93.
[13] E. Kalaitzoglou, T.M. Griffin, M.B. Humphrey. “Innate immune responses and osteoarthritis”. Current Rheumatology Reports. 2017, vol. 19(8), pp. 1-6.
[14] L.C. Chan, H.H. Li, P.K. Chan, C. Wen. “A machine learning-based approach to decipher multi-etiology of knee osteoarthritis onset and deterioration”. Osteoarthritis and Cartilage Open. 2021, vol. 3(1), pp. 100-135.
[15] C. Kokkotis, S. Moustakidis, G. Giakas, D. Tsaopoulos. “Identification of risk factors and machine learning-based prediction models for knee osteoarthritis patients”. Applied Sciences. 2020, vol. 28,10(19), pp. 6797.
[16] A.J. Aljaaf, A.J. Hussain, P. Fergus, A. Przybyla, G.J. Barton. “Evaluation of machine learning methods to predict knee loading from the

978-1-6654-5432-2/22/$31.00 ©2022 IEEE

https://www.cdc.gov/chronicdisease/resources/publications/factsheets/arthritis.htm
[4] J. Maiuolo, F. Oppedisano, S. Gratteri, C. Muscoli, V. Mollace. “Regulation of uric acid metabolism and excretion”. International journal of cardiology. 2016, pp. 213:8-14.
[5] I. Widyadharma, P.R. Dewi, I.A. Wijayanti, D.K. Utami. “Pain related viral infections: A literature review”. The Egyptian journal of neurology, psychiatry and neurosurgery. 2020, vol. 56(1), pp. 1-6.
[6] A. Alexos, C. Kokkotis, S. Moustakidis, E. Papageorgiou, D. Tsaopoulos. “Prediction of pain in knee osteoarthritis patients using machine learning: Data from Osteoarthritis Initiative”. In2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA 2020 Jul 15 (pp. 1-7). IEEE. movement of body segments”. In2016 International Joint Conference on Neural Networks (IJCNN) 2016 Jul 24 (pp. 5168-5173). IEEE.
[17] J. Bianchi, A.C. de Oliveira Ruellas, J.R. Gonçalves, B. Paniagua, J.C. Prieto, M. Styner, T. Li, J. Zhu, J. Sugai, W. Giannobile, E. Benavides. “Osteoarthritis of the Temporomandibular Joint can be diagnosed earlier using biomarkers and machine learning”. Scientific reports. 2020, vol. 10(1), pp. 1-4.