AI4AFRICA 10XDS PROGRAM

TRAINING THE NEXT GENERATION OF AFRICAN DATA SCIENTISTS

Knee Osteoarthritis: Modifying the Cellular Therapy Protocol

Authors:
Marwa Adam
Olusegun Emmanuel Ajibola
Kerol DJOUMESSI

Date: January 22, 2022

Contents

1 Introduction

Arthritis affects joints in the hands, knees, hips and spine. It affects millions of people worldwide. Knee Osteoarthritis (KOA) is the most common cause of arthritis that occurs when the protective cartilage that cushions the ends of the bones wears down over time. It affects the ageing population of the world mostly individuals between the age range of 40-85 years old. Unfortunately, it can not be treated using antibiotics since the joints have no veins or arteries.

Osteoarthritis can make one disabled since it causes difficulties in movements and affects one's daily life. People with Osteoarthritis suffer from pain, stiffness and in the worst case they usually have a poor quality of life.

Possible treatments for knee osteoarthritis usually include clinical operations in the infected knee. Currently, there has been another alternative which is a cellular therapy protocol. The cellular therapy protocol consists of a number of injections (stages), normally four injections. The period between injections is three (3) weeks.

The treatment will not change the patient's life totally, but it minimizes pain and as a result, you may minimize the number of pain killers and/or non-steroidal anti-inflammatory drugs (or NSAIDs) being taken. At times, it may eliminate the need to take these drugs.

At the Al-Neelain Stem Cells Center (ASCC), Al-Neelain University, Sudan, they conducted the treatment using the cellular therapy protocol, which is the first time this method of treatment will be performed in Sudan. Nonetheless, there is a need to configure the therapy to ascertain the progress made after each injection and to check if the four (4) series of injections are enough or not. Moreover, to determine the number of doses needed for each patient depending on the initial status of the patient and the improvement(s) they have made after the treatment.

2 Knee Injury and Osteoarthritis Outcome Score (KOOS)

2.1 Definition

The KOOS is a knee-specific instrument, developed to assess the patients' opinion about their knee and associated problems. It evaluates both short-term and long-term consequences of a knee injury.

To measure how damaged a knee is, the patient is required to fill a questionnaire. The responses in the questionnaire are evaluated to derive the Knee Injury and Osteoarthritis Outcome Score (KOOS). The questionnaire consists of five (5) different subscales which

Subscale	Number of Questions	Required Answers
Pain	9	5
Symptoms	7	4
ADL	17	9
Sport	5	3
QOL	4	2

Table 1: Subscale questions and the minimum number of required answers (ref).

are: **symptoms, pain, physical function in daily life, physical function while doing sports** and **the quality of life**. Moreover, each subscale includes some questions, and each question has five different answers given a score ranging from 0 to 4 (see figure 1).

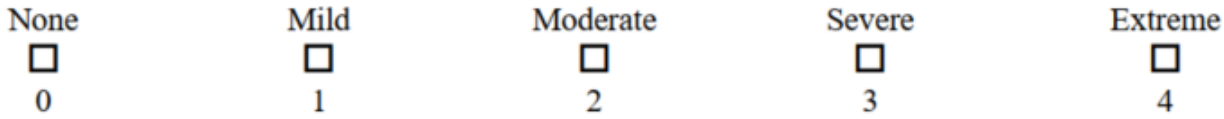

Figure 1: Answers together with the grades

To deal with the problem of missing values that most often arises when calculating the KOOS, a change in how to manage missing items was introduced in 2012. Indeed, two missing items were allowed in each subscale previously. But from 2012, according to the new rule for missing items, at least 50% of the items should be responded to. As result, the number of questions per subscale together with the allowed number of answers needed for calculation of each subscale score is determined in table 1.

2.2 KOOS computation

The KOOS is computed by averaging the sub-KOOS resulting from each subscale. Typically, each subscale is computed by applying the mean of the observed items within the subscale, divide by 4, and multiply by 100. Then, the resulting number is subtracted from 100. As result, the formulas for calculating each subscale are as follows:

$$\text{Pain KOOS} = 100 - \frac{\text{Mean score}\,(P_1 - P_9) \times 100}{4}$$

$$\text{Symptoms KOOS} = 100 - \frac{\text{Mean score}\,(S_1 - S_7) \times 100}{4}$$

$$\text{ADL KOOS} = 100 - \frac{\text{Mean score}\,(A_1 - A_{17}) \times 100}{4}$$

$$\text{Sport KOOS} = 100 - \frac{\text{Mean score}\,(SP_1 - SP_5) \times 100}{4}$$

$$\text{QOL KOOS} = 100 - \frac{\text{Mean score}\,(Q_1 - Q_4) \times 100}{4}$$

As a result, we have the KOOS subscale estimate for that particular cross-sectional assessment of the individual patient. The final KOOS of the considered patient will be the mean of those sub-KOOS.

3 Data

3.1 Data Collection

The dataset was collected at the Al-Neelain Stem Cells Center (ASCC) as a part of the treatment procedure. It includes two separate data files, the first file (see figure 2) include the key information about the patients, such as age, gender, which knee is treated, whether the patient has a chronic disease or not and the haematological blood test results (CBC). The second data file (see figure 3) contains the questionnaire answers for each patient, this consists of four (4) questionnaires per patient.

ID	Age	Gender	Chronic Disease	Knee	WBC	RBCs	HGB	HCT	MCV	MCH	MCHC	PLTs	...
0	46	Female	None	Left	5.41	4.8	10.6	33.9	38.1	26.1	31.4	210.3	...

Figure 2: Patient information

	ID	knee	S1	S2	S3	S4	S5	S6	S7	P1	...	A17	SP1	SP2	SP3	SP4	SP5
0	1.0	left	Never	sometimes	Never	Always	Always	None	None	Always	...	Mild	Extreme	Extreme	Extreme	Extreme	extreme
1	NaN	NaN	Never	rarely	rarely	sometimes	sometimes	None	None	daily	...	Mild	Extreme	Extreme	Extreme	Extreme	extreme
2	NaN	NaN	sometimes	sometimes	sometimes	sometimes	sometimes	Moderate	moderate	daily	...	Mild	Extreme	Extreme	Extreme	Severe	Extreme
3	NaN	NaN	rarely	sometimes	sometimes	sometimes	sometimes	Moderate	Mild	daily	...	Mild	Extreme	Extreme	Extreme	Severe	Extreme

Figure 3: Questionnaire

The dataset contains 75 patients (25 male and 50 female) who completed the treatment for either their left or right knee. Note that patients who had the treatment for both knees were considered as a different case. Each patient had 4 different questionnaires, in total we had 75×4 data points or rows in the questionnaire file.

3.2 Patient follow-up process

At the first time a patient comes to the center, the information collected are the main information, also, (s)he will be asked to fill his/her first questionnaire. Then the first KOOS will be calculated for the first questionnaire. In the same day, the patient will

have his first injection of the treatment. The second injection will be after three weeks, and the patient will be asked to fill the second questionnaire before taking the injection. Note that the questionnaires are the same every time.

The third questionnaire will be filled after three weeks from taking the second injection, which is before taking the third injection. The very last injection is optional, it will be after three weeks from the third, and the questionnaire will be filled just before the injection.

4 Data Analysis and prediction model

4.1 Objectives

As a first step of the research project, the mains objectives are

- Digitize the KOOS computation process
- Predict the recovery rate before the next injection

4.2 Data Analysis

The first task was to calculate the KOOS. After then, we determine the improvement that each patient has made after taking each injection, furthermore, after completing the whole treatment. We named this improvement ***recovery rate*** which is simply the difference between two KOOS scores. The figure 4 summarize the process we followed to compute the KOOS and the recovery rate.

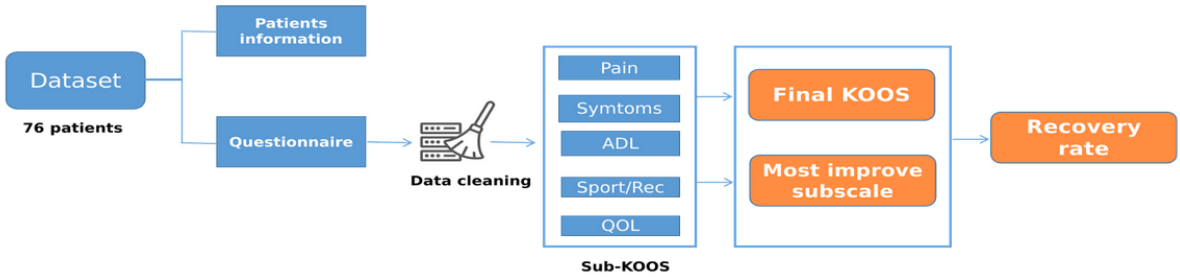

Figure 4: KOOS computation pipeline

Calculating the KOOS allowed us to be able to evaluate the improvement for the patients after each of the injections, also, after completing the treatment. The Recovery rate has been is calculating by taking the differences between the KOOS as follows:

$$RR_1 = 2nd\ KOOS - 1st\ KOOS$$

$$RR_2 = 3rd\ KOOS - 2nd\ KOOS$$
$$RR_3 = 4th\ KOOS - 3rd\ KOOS$$

where

- $1st\ KOOS$ denotes the first KOOS before the first injection,
- $2nd\ KOOS$ denotes the second KOOS (the one after the first injection and before the second injection),
- $3rd\ KOOS$ denotes the third KOOS (the one after the 2nd injection and before the 3rd injection),
- $4th\ KOOS$ denotes the fouth KOOS (the one after the 3rd injection and before the 4th injection).

Note that The overall recovery rate can be obtained either by summing up the three recovery rates, or by subtracting the first KOOS score from the last KOOS score.

4.3 Machine learning model

The goal of the prediction or machine learning model is to predict the treatment outcome (the recovery rate) for the patient before the ending the treatment, given the main information together with the KOOS and the recovery rate we already have.

For example, after the first injection, we already have the first KOOS and the machine learning task will be to predict the recovery rate before the next injection. So, we build three different model and depending on the information (KOOS and recovery rate) we have about a patient, we will use one of the three models to predict the recovery rate.

On the first model, we assume that we have the first KOOS and would like to predict what will be the first recovery rate. The second model assumes that we have the first two KOOS and want to predict the recovery rate before the third injection. The third model assume we have the first three KOOS and would like to predict the final recovery rate.

4.3.1 Machine learning pipeline

Figure 6 shows an overview of the machine pipeline where the inputs consist of questionnaires, key patient information, available KOOS, and recovery rate. The output is a continuous value indicating the recovery rate. Since the output is continuous, the prediction is made by finding the parameters of a linear regression model.

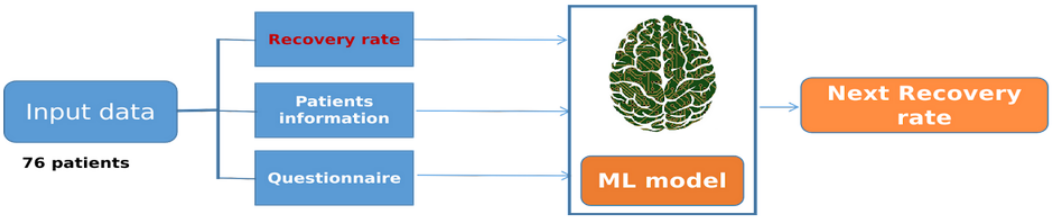

Figure 5: Machine learning pipeline

4.3.2 Model

We tested two different models: the **xgboost regressor**, and the **lightgbm regressor**. We then selected the best model based on a cross validation process with the same hyperparmeters.
In addition, we use 80% or 60 data point of our dataset for training the model and the rest (20% or 15 data point) for evaluation. To select the best parameters, we used the **Randomized search** provided in *sklearn*

5 Results

The most improved subscale

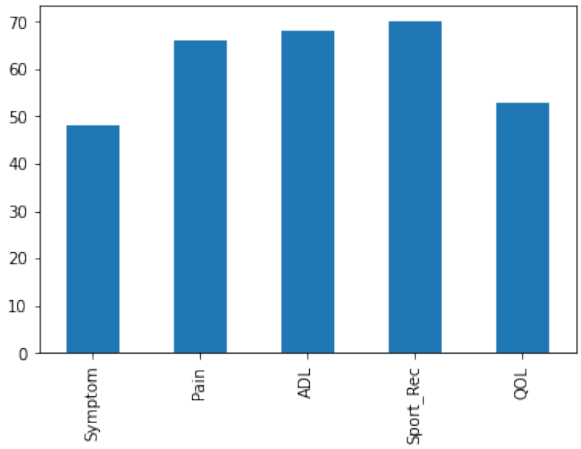

Figure 6: Most imrprove subscale

6 Conclusion

References