Sigma J Eng & Nat Sci 38 (2), 2020, 1051-1059

Sigma Journal of Engineering and Natural Sciences
Sigma Mühendislik ve Fen Bilimleri Dergisi

Research Article

PREDICTION OF DEMOGRAPHICAL CHARACTERISTICS USING K-MEANS ALGORITHMS

Murat SARI*[1], Can TUNA[2], Ibrahim DEMIR[3]

[1]*Department of Mathematics, Yildiz Technical University, ISTANBUL;* ORCID: 0000-0003-0508-2917
[2]*Department of Mathematics, Yildiz Technical University, ISTANBUL;* ORCID: 0000-0002-1459-8782
[3]*Department of Statistics, Yildiz Technical University, ISTANBUL;* ORCID: 0000-0002-2734-4116

Received: 05.11.2019 Revised: 10.03.2020 Accepted: 12.05.2020

ABSTRACT

It is crucially important to predict demographic characteristics of criminals from the footprint area at the crime scene. Demographic characteristics include age, weight, height and gender. This article has thus investigated the effect of the tibial rotations on predictions of the demographical characteristics using the K-Means (KM) clustering algorithms. Satisfactorily important predictions have been carried out through the dataset consisting of 484 healthy subjects in the designed study here. The produced results revealed that it is of great potentiality to do also for criminals. The results are therefore believed to be vitally important for most fields of forensic science. Specifically, it can provide important clues when diagnosing criminals. Note that the KM algorithms have been found to be very encouraging processing system for modelling in the assessment of the demographic characteristics.
Keywords: Forensics science, tibial rotation, biomechanics, KM algorithm.

1. INTRODUCTION

One of the most interesting issues of contemporary science is to develop methods and models for studying various systems. In this way, the knowledge is presented in terms that can be used. Progressively, the mathematics turns into a language of innovations and smart technologies in the industrial society. Thus, most problems encountered in nature are usually represented by mathematical models. To analyse those problems emerged in broad range of science, mathematical modelling has been considered as an important tool. Advent of computers, producing algorithms and progress in computer programming have made life easier in solving such intricate problems of science. This is also the case in problems arisen in biomechanics. To make the best biomechanical decisions, medical prediction plays a crucial role for health providers. Notice that demographical prediction can provide important clues when diagnosing criminals. Many researchers have particularly focused on analysis of the knee motion and many methods have been suggested in dealing with the tibial motion [1,2].

As underlined in the literature [3,4] the knee joint is very complex joints in the musculoskeletal system. To assess the motion of the knee joint, various approaches have been

* Corresponding Author: e-mail: sarim@yildiz.edu.tr, tel: (212) 383 43 60

designed to describe the motion of the knee [5-10]. Due to external rotation related to knee extension, excess ive internal rotation during the stance phase of walking can postpone the natural external rotation while the knee extends. Therefore, some researchers [4,8,10] concentrated on study of the tibial motion involving the internal and external rotations.

Assessment of the tibial motion is of great difficulty for biomechanical points of view. Even though it is encountered striking studies in the literature, to the best knowledge of the authors, effects of the tibial rotations on demographical prediction have not been analysed yet. To tackle this study, the KM clustering algorithm has been introduced. By using various techniques [4,5,10-16], most studies have usually focused on specifying the tibial motion through physical information rather than investigating the effect of the tibial motion on the demographic characteristics. Since demographic characteristics of an offender plays a vitally important role in forensic research, this article has investigated the effect of the tibial rotations on demographical parameters using the KM clustering algorithm. There are several methods like using statistics at the modelling stage. The algorithm which are being used rather frequently in the last decade provide an alternative model for many clustering procedures. A great deal of clustering algorithms and thus the KM clustering algorithm have been encountered in the literature. Although various versions of the KM algorithm were effectively used in different fields of science [17], to the best knowledge of the authors, the KM clustering algorithm has not been implemented to predict the demographical characteristics based on the tibial motion so far. Since the KM clustering algorithm is more flexible, it has been used for the trustworthy data processing and consequent interpretations in the current work. The KM algorithm, as a general modelling instrument, makes the estimation process possible for many different patterns.

The rest of the paper is organized as follows. Section 2 is devoted to the material and method. Section 3 exhibits discussion and analysis of the produced results. Finally, Section 4 presents final remarks and recommendation.

2. MATERIAL AND METHOD

2.1. Participants and study design

The data for 484 healthy subjects were provided from the work of Cetiner and Sari [10]. The subjects do not have any health problems or knee joint injuries. The data includes assessment of age, weight, height information and gender of 484 volunteers. The physical parameters (age, weight, height and gender) of each subject are indicated in Figures 1-3.

In the data; tibial rotation values of each subject were given as right tibial external rotation (RTER), right tibial internal rotation (RTIR), left tibial external rotation (LTER), left tibial internal rotation (LTIR) as seen in Figure 2. Totally, for every subject, it was 8 parameters as 4 of them are tibial rotation values (RTER, RTIR, LTER, LTIR) and 4 of them are physical factors (age, weight, height). The rotation values are input and the physical parameters are output.

All subjects considered in the present study had no musculoskeletal, neurological or systemic disease that might influence testing. To see effect of the tibial rotations on demographical characteristics (weight, height, age, gender; N: 484, Female: 238, Male: 246), individuals who hold the inclusion criteria were studied. The classification of the data is given by taking into account the BMI obesity ranges of the WHO (see Table 1).

Table 1. Body mass index (BMI) values and obesity interval for the data

BMI	BMI	Subjects		
Class	Interval	Female	Male	Total
Thinness	0 –18.5	37	4	41
Normal	18.5 – 25	178	203	381
Obese	25+	23	39	62

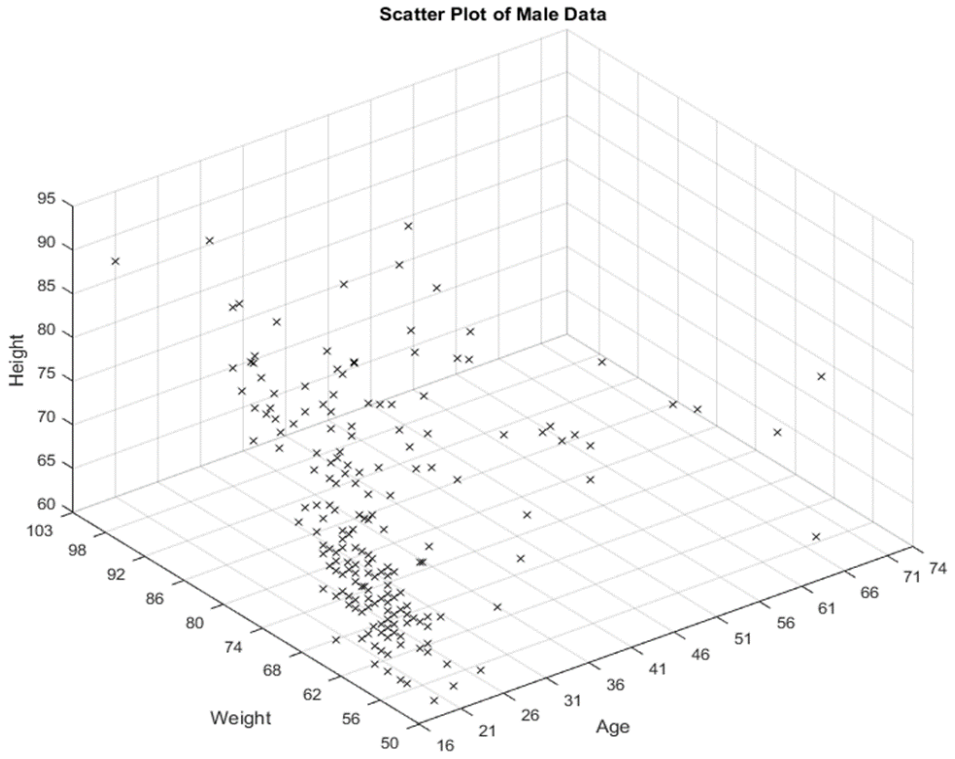

Figure 1. Scatter plot of the male data consisted of age, weight and height parameters.

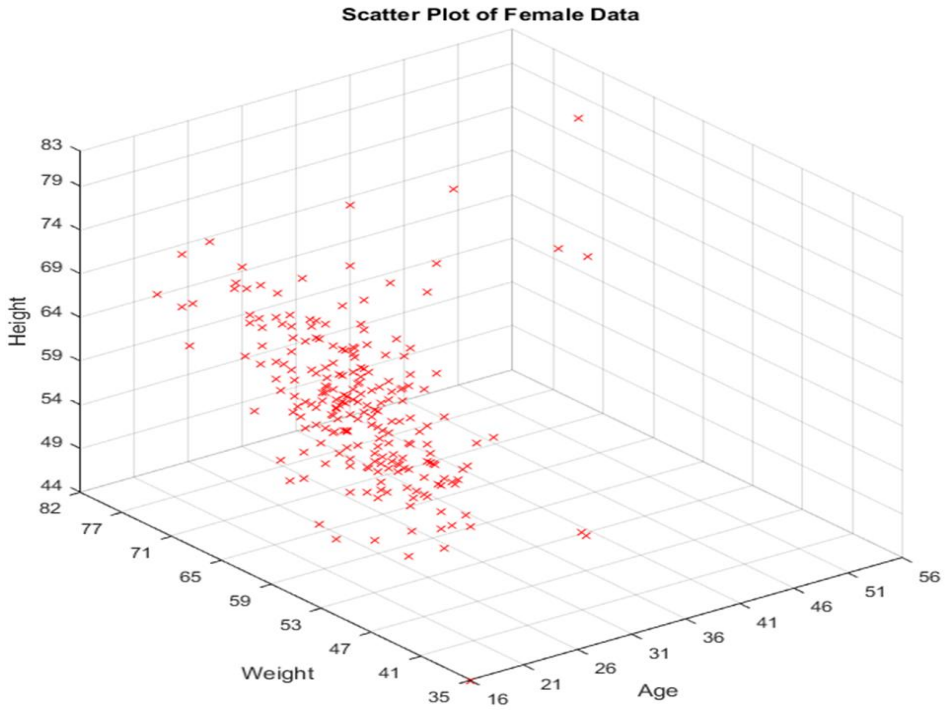

Figure 2. Scatter plot of the female data consisted of age, weight and height parameters.

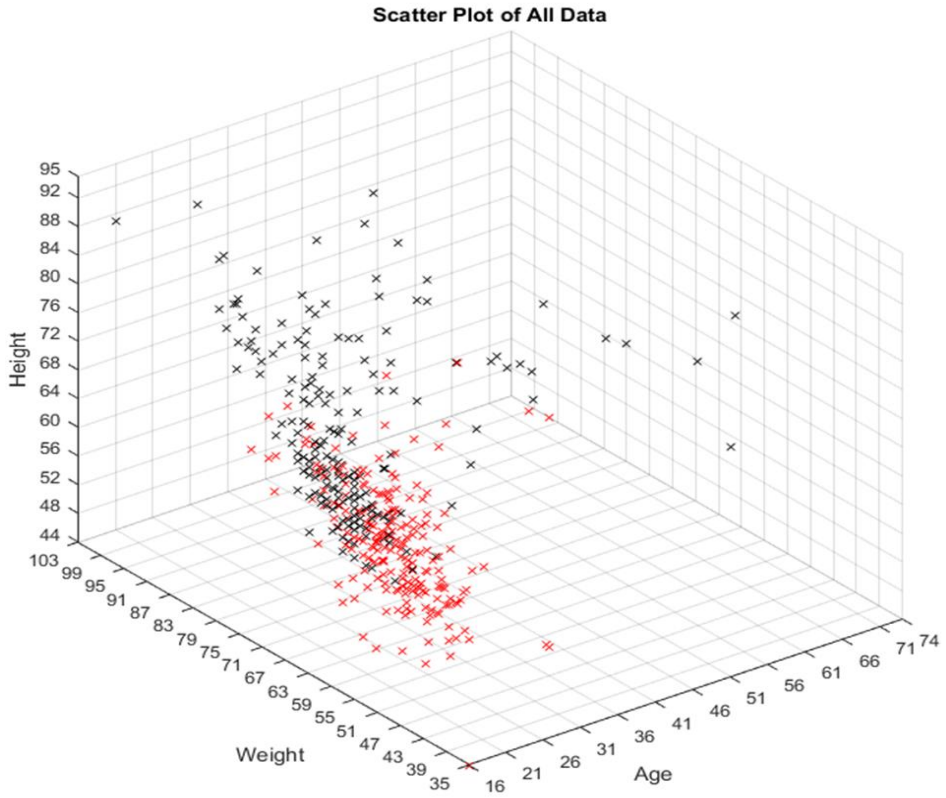

Figure 3. Scatter plot of all data consisted of age, weight and height parameters.

2.2. K-Means Algorithm

K-Means (KM) clustering was, for the first time, explored by MacQueen [18]. The KM clustering algorithm is one of the fastest, simplest and the most commonly used one among clustering methods for the problems of interest. The algorithm leads N sample data to categorize K clusters according to their distance to each other. The KM algorithm aims at finding the best cluster centre for each iteration. Centre of the clusters are updated for each iteration. The algorithm must have an objective function and the corresponding function is to be minimization problem. The distance is computed using the algorithm as, in the Euclidean sense [19,20].

$$D = \sum_{j=1}^{K} \sum_{i=1}^{N} \left\| x_i - C_j \right\|^2 \quad (1)$$

where x_i, $1 \le i \le N$, and C_j, $1 \le j \le K$, indicate the set of N data and set of cluster centroids, respectively. The distance between any two p-dimensional points X_i and X_j is calculated by Das et al. [21]

$$d(X_i, X_j) = \sqrt{\sum_{m=1}^{p} (X_{im} - X_{jm})^2} \quad (2)$$

The flow chart of the KM algorithm has been exhibited in Figure 4. The pseudocode of the KM algorithm has been presented as shown in Pseudocode 1.

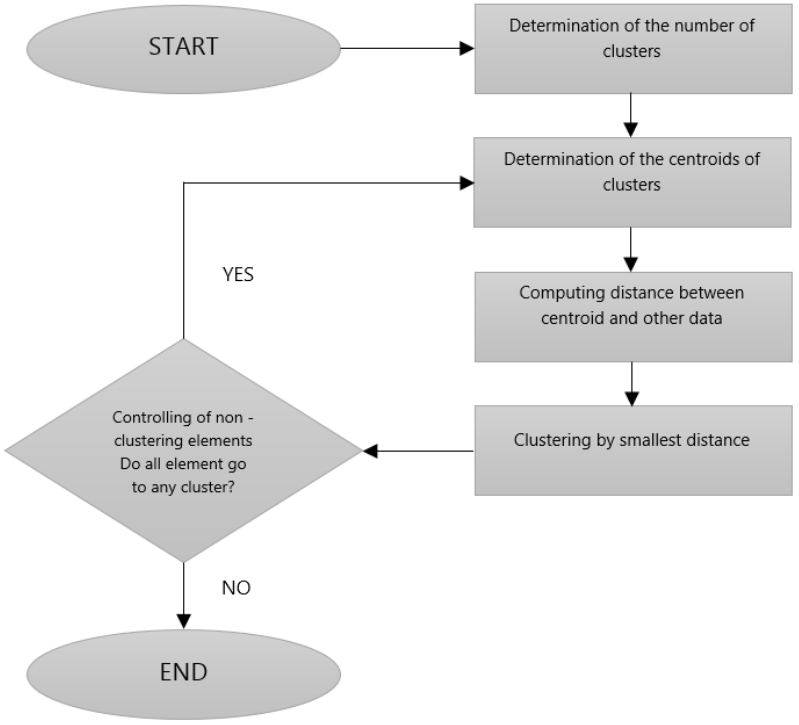

Figure 4. Flow diagram of the K-Means clustering algorithm

```
begin
  Determining the number of centroids (k centroids)
  Randomly selecting k centroids from population
  repeat
    for i to population
      Computing distance between each centroid and other data
      Clustering by smallest distance for each data
    end
  until There is no non-clustering data
end
```

Pseudocode 1. Pseudocode of the KM clustering algorithm

3. DISCUSSION AND ANALYSIS

Since the day that they were found, the KM clustering have been used as very successful algorithm in different disciplines. In this study, the algorithm has been applied to the prediction of the demographical characteristics based on the tibial rotations for the first time. It was tested if it

would be successful in this area as is the case other areas of science. The proposed algorithm has produced very encouraging results in the prediction through the tibial rotation types (see Tables 2-6). The application in this field is strongly believed to be an important milestone in forensic search and to help professional workers in predicting the characteristic factors of individuals/criminals through their tibial rotations.

The proposed algorithm works much better when relatively enough data is considered. Since the data consists of subjects mostly younger than 30 years old, this work may not be very decisive enough for that subjects who are older than 30 years old in this data.

Table 2. Comparison of the results produced by the KM algorithm with the real values

		Real		KM Algorithm	
Case	Inputs	Female	Male	Female	Male
1	RTER RTIR LTER LTIR	238	246	208	205
2	LTER LTIR	238	246	193	208
3	RTER LTER	238	246	218	172
4	RTER RTIR	238	246	209	191
5	RTIR LTIR	238	246	159	191

Table 3. Comparison of the KM results with the real gender data (the input variables: RTER, RTIR, LTER, LTIR)

		K-Means Results		
	Classification	Female	Male	Total
Real Values	Female	208	41	249
	Male	30	205	235
	Total	238	246	484
Accuracy	85.33%			

Table 4. Comparison of the results of the KM algorithm with the real obesity data for female subjects (the input variables: RTER, RTIR, LTER, LTIR)

		K-Means Results			
	Classification	Thinness	Normal	Obese	Total
Real Values	Thinness	23	57	0	80
	Normal	10	91	8	109
	Obese	4	30	15	49
	Total	37	178	23	238
Accuracy	54.20%				

Table 5. Comparison of the results of the KM algorithm with the real obesity data for male subjects (the input variables: RTER, RTIR, LTER, LTIR)

		K-Means Results			
	Classification	Thinness	Normal	Obese	Total
Real Values	Thinness	2	30	1	33
	Normal	2	108	15	125
	Obese	0	65	23	88
	Total	4	203	39	246
Accuracy	54.07%				

Table 6. Comparison of the results of the KM algorithm with the real obesity data for all subjects (the input variables: RTER, RTIR, LTER, LTIR)

	Thinness	Normal	Obese
Real	41	381	62
KM Algorithm	28	172	40
Rate	68.29%	45.14%	64.52%

4. CONCLUSIONS AND RECOMMENDATION

Knowing demographic characteristics of individuals is a vital issue, and thus this paper has concentrated on the prediction of the characteristics from a population through various tibial rotations using the KM algorithm. The results of the algorithm for a population revealed that demographic characteristics (gender and obesity level) of the subjects can be estimated from their tibial rotations. It has been concluded that findings are expected to be very useful in giving directions to perform their occupations. The algorithm has been found to be very encouraging processing system for modelling in the assessment of the demographic characteristics. This study encourages tremendously us to specifically design a research for a dataset of subjects. Additionally, in the future, a more systematic data might be more useful for the prediction. In the forthcoming works, more detailed and thus more illustrative results can be produced with various datasets.

Acknowledgements

The authors would like to thank anonymous reviewers for their valuable comments and suggestions to improve the paper.

REFERENCES

[1] J.C. Kupper, B. Loitz-Ramage, D.T. Corr, D.A. Hart, J. L. Ronsky, Measuring knee joint laxity: a review of applicable models and the need for new approaches to minimize variability, *Clinical Biomechanics*, 22(1), 1-13, 2007.

[2] J. O. S. E. P. H. Hamill, B. T. Bates, and K. G. Holt, Timing of lower extremity joint actions during treadmill running, Medicine and Science in Sports and Exercise, 24(7), 807-813, 1992.

[3] S.F. Dye, An evolutionary perspective of the knee, *J Bone Joint Surg Am*, 69(7), 976-983, 1987.

[4] M. Sari, B.G. Cetiner, Predicting effect of physical factors on tibial motion using artificial neural networks, *Expert Systems with Applications*, 36(6), 9743-9746, 2009.

[5] L. R. Osternig, B. T. Bates, and S. L. James, "Patterns of tibial rotary torque in knees of healthy subjects," Medicine and Science in Sports and Exercise, vol. 12, no. 3, pp. 195-199, 1980.

[6] O. S. Mills and M. L. Hull, Rotational flexibility of the human knee due to varus/valgus and axial moments in vivo, Journal of Biomechanics, 24(8), 673-690, 1991.

[7] K.E. Moglo, A. Shirazi-Adl, Biomechanics of passive knee joint in drawer: load transmission in intact and ACL-deficient joints, *The Knee*, vol. 10, no. 3, pp. 265-276, 2003.

[8] T. Armour, L. Forwell, R. Litchfield, A. Kirkley, N. Amendola, P. J. Fowler, Isokinetic evaluation of internal/external tibial rotation strength after the use of hamstring tendons for anterior cruciate ligament reconstruction, The American Journal of Sports Medicine, 32(7), 1639-1643, 2004.

[9] P. Johal, A. Williams, P. Wragg, D. Hunt, W. Gedroyc, Tibio-femoral movement in the living knee. A study of weight bearing and non-weight bearing knee kinematics using 'interventional' MRI, Journal of Biomechanics, 38(2), 269-276, 2005.

[10] B.G. Cetiner, M. Sari, Tibial rotation assessment using artificial neural networks, Mathematical and Computational Applications, 15(1), 34-44, 2010.

[11] M. Sari, Relationship between physical factors and tibial motion in healthy subjects: 2D and 3D analysis, Advances in Therapy, 24(4), 772-783, 2007.

[12] A. Cimbiz, U. Cavlak, M. Sari, H. Hallaceli, F. Beydemir, A new clinical design measuring the vertical axial rotation through tibial shaft resulting from passive knee and subtalar joints rotation in healthy subjects: a reliability study, Journal of Medical Sciences, 6(5), 751-757, 2006.

[13] D. Tiberio, The effect of excessive subtalar joint pronation on patellofemoral mechanics: a theoretical model. Journal of Orthopaedic & Sports Physical Therapy. 1987;9(4):160-165.

[14] L. Lang, R. Volpe, Measurement of tibial torsion. Journal of the American Podiatric Medical Association. 1998; 88(4):160-165.

[15] M.A. Freeman, V. Pinskerova, The movement of the knee studied by magnetic resonance imaging. Clinical Orthopaedics and Related Research, (2003): 410, 35–43.

[16] F. Lin, G. Wang, J.L. Koh, R.W. Hendrix, L.Q. Zhang, In vivo and noninvasive three-dimensional patellar tracking induced by individual heads of quadriceps. Medicine and Science in Sports and Exercise, 2004, 36(1), 93–101.

[17] H. Yu, Z. Liu, Wang, G. An automatic method to determine the number of clusters using decision-theoretic rough set. International Journal of Approximate Reasoning 2014, 55, 101–115.

[18] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. vol. 1. Oakland, CA, USA. 1967. p. 281-297.

[19] A.K. Jain, Data clustering: 50 years beyond K-means. Pattern recognition letters. 2010; 31(8):651-666.

[20] A. Alguwaizani, Degeneracy on K-means clustering. Electronic Notes in Discrete Mathematics. 2012; 39:13-20.

[21] S. Das, A. Abraham, A. Konar, Automatic clustering using an improved differential evolution algorithm. IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum. 2008; 38:218–237.