

Artificial Intelligence enhanced kinematic analysis of basketball free throws

BEKIR KARLIK¹, MUSA HAWAMDAH²

¹Department of Computer Engineering, Epoka University, Tirana, ALBANIA

²Department of Computer Engineering, Selcuk University, Konya, TURKEY

Published online: December 30, 2024

Accepted for publication: December 15, 2024

DOI:10.7752/jpes.2024.12321

Abstract:

Problem Statement and Approach: In basketball, the success of free throws is important in terms of the ball's exit angle, its proper position in the air and its optimum speed kinematic features. This study investigates the kinematic characteristics of basketball players executing free throws both before and after experiencing fatigue, using artificial Intelligence (AI). **Materials and Methods:** We used various supervised machine learning algorithms including: k-Nearest Neighbours (k-NN), Naïve Bayes, Support Vector Machines (SVM), Artificial Neural Networks (ANN), Linear Discriminant Analysis (LDA), and Decision Tree. These algorithms were used for classifying features derived from motion data collected from players to uncover patterns and variations in their shooting mechanics under different levels of fatigue. The elbow, trunk, knee and ankle joint angles are measured at the ball release point when the players make successful and unsuccessful shots before and after fatigue. There are two used methods for classification of these features: the first one is used directly row data; the other is used reduced data by using principal component analysis (PCA). For both, data normalized between 0-1 before applying to the classifier. **Results:** We obtained the best classification accuracy as 98.44% by using Naïve Bayes classifier for row data. Moreover, the results of ANN with reduction data by using PCA have shown the best classification accuracy as 95.31%. The findings reveal distinct patterns and variations in shooting mechanics induced by fatigue and highlight the effectiveness of machine learning models in analysing biomechanical data. **Discussion and conclusion:** These results can aid in developing training programs to enhance performance and consistency in fatigued states. This research underscores the potential of AI and data-driven approaches in sports biomechanics to uncover valuable insights into athlete performance and fatigue management.

Keywords: Intelligent Algorithms, Sports Biomechanics, Kinematic Data, Fatigue-Induced Variations

Introduction

The differences between functional skills and skills based-movement patterns have been observed in studies conducted across various sports. Evaluating functional skills is more challenging than evaluating skill-based movement patterns (Goktepe et al., 2009; Abdelkerim et al., 2007; Chappell et al., 2005). For instance, Goktepe et al. (2009) utilized statistical analysis to demonstrate the influence of ankle, shoulder, and elbow angles in tennis serves. Abdelkerim et al. (2007) presented computerized time-motion analyses for basketball players, while Chappell et al. (2005) examined altered motor control strategies in landing and jumping maneuvers during three stop-jump tasks, performed pre- and post-fatigue exercises. Evaluating factors such as skill-based contraction, appropriate muscle firing timing, and joint positioning is comparatively easier. Notably, differences exist between movement execution and skill acquisition among individuals. In basketball free throws, joint angles are fundamental features sufficient to classify shots into various categories (Schmidt et al., 2012; Ge, 2024; Zhang & Chen, 2024).

Fatigue, a natural consequence of human activity, impacts both cognitive and learning capacities in athletes during training and competition. While most studies identify fatigue as a critical factor influencing performance (Forestier & Nougier, 1998; Aprianono et al., 2006), some indicate no effect of fatigue on free throw performance in basketball (Uygur et al., 2010; Rusdiana et al., 2019; Li, 2021; Bourdas et al., 2024). For example, Uygur et al. (2010) observed no significant effects of fatigue on free throws based on statistical kinematic analysis. Similarly, Rusdiana et al. (2019) analyzed free throw kinematic movements using SPSS, while Bourdas et al. (2024) focused on the impact of fatigue on 3-point jump shots. Li et al. (2021) examined the kinematic effects of fatigue on female basketball players' shooting performance. All these studies employed statistical methods; no AI or soft computing techniques have been found in the literature to analyze basketball free throw kinematics.

Efficient data analysis in recent decades has significantly improved productivity across various domains using soft computing methods. However, most studies in sports science have focused on specific games to explore the roles of different data sources or machine learning techniques in structural analysis and semantic extraction. This study is the first to apply machine learning methods to the kinematic analysis

of basketball free throws. Various supervised machine learning techniques are used to assess the impact of fatigue on player performance. As shown in Figure 1, data on joint angles (elbow, trunk, knee, and ankle) at the ball release point were collected before and after fatigue (Uygur et al., 2010). Results show that Naïve Bayes yielded the highest classification accuracy (98.44%) with raw data, while ANN performed best (95.31%) with PCA-reduced data. The effect of feature reduction on classification accuracy was also analyzed.

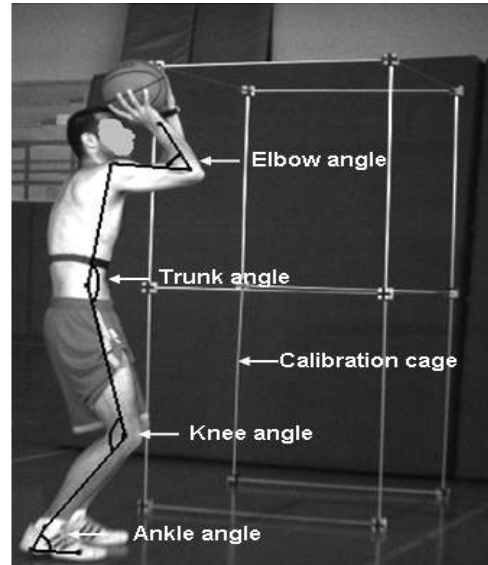


Figure 1. Joints angles elbow, trunk, knee and ankle for expert basketball player [Uygur,2010]

Materials and Methods

Machine learning is used and serves to decision-making, which faces uncertainty, incomplete knowledge, problems, and complex datasets, and the imperfection of the heterogeneous multiple participants (e.g., decision- makers, diagnosis classifiers, etc.). However, machine learning is used to analyze the kinematics of players in many kinds of sports and human activities as an effective method (Fukuchi et al., 2011; Smeeton et al., 2004; Begg et al., 2005). Fukuchi et al. (2011) proposed a support vector machine (SVM) method for detection of age-related changes in running kinematics. Smeeton et al. (2004) have presented if skilled and less skilled players could transfer pattern recognition skills across similar and dissimilar sports. Begg et al. (2005) have proposed different types of SVM models that are able to map the underlying data structure relating to young and aging populations and discriminate between young and elderly kinematics in walking.

A learning algorithm can modify current and future instructions with prior knowledge to improve performance. The best bet for choosing a machine learning algorithm for a classification problem is to test several different algorithms and select the best one through validation. Also, in measurement-based studies, it may be necessary to apply appropriate data reduction techniques to eliminate erroneous measurements.

Data description

The data were collected in a previous study from eight basketball players before and after fatigue, categorized into four groups: (pre-successful, pre-unsuccessful, post-successful and post-unsuccessful) (Uygur et al., 2010). Each player attempted eight free throw shots: two successful and two unsuccessful shots before fatigue, as well as two successful and two unsuccessful shots after fatigue. Consequently, the dataset comprises 64 samples, with four input features for classification and four corresponding outputs. The outputs indicate whether the shot was successful or unsuccessful, both before and after fatigue. Each sample includes four features which are; the elbow, trunk, knee, and ankle joint angles at the ball release point, as summarized in Table 1.

Table 1. Data description.

Shots	Pre-fatigue	Post-fatigue
Successful	16	16
Unsuccessful	16	16
Total	32	32

Data Normalization

Data normalization enhances modeling and prevents numerical complexity. Various normalization methods exist, including Min-Max, outlier assessment, histogram trend analysis, and Exponential Moving Average. Min-Max normalization is particularly effective as a linear scaling algorithm. It transforms the original dataset into a new range, typically between 0 and 1, which accelerates the learning process in

machine learning applications (Jain and Bhandare, 2011). Additionally, it preserves all relationships among data values. In this study, Min-Max normalization was applied, and the data were normalized between 0 and 1. The differences between the data before and after normalization are presented in Table 2.

Table 2. The differences between data before and after normalization

Features	Before-Normalization	After-Normalization
	Average	Average
Elbow-angle	97.141 +/- 16.827	0.453 +/- 0.181
Trunk-angle	178.125 +/- 9.221	0.504 +/- 0.266
Knee-angle	159.625 +/- 11.598	0.647 +/- 0.186
Ankle-angle	71.281 +/- 6.973	0.452 +/- 0.273

Data Reduction with Principal Component Analysis

Principal component analysis (PCA) is used to analyze data, recognize patterns, and reduce dimensionality in a dataset while minimizing information loss (Karlik et al., 2021). PCA can also reduce parameter estimation errors and the computational costs by reducing the number of dimensions in the feature space and extracting a subspace that defines the data effectively. In this study, based on cumulative variance, when the variance threshold is set at 95%, principal components selected are sufficient to explain most of the variance, as shown in Table 3. Consequently, we reduced the number of features from four to two using PCA.

Table 3. Principal component analysis, SD, Proportion of Variance and Cumulative variance when the variance threshold is 95%

Component	SD	Proportion of Variance	Cumulative variance
PC 1	0.323	0.490	0.490
PC 2	0.223	0.233	0.723
PC 3	0.198	0.185	0.908
PC 4	0.140	0.092	1.000

Used Machine Learning Algorithms

Machine learning, a subset of Artificial Intelligence (AI), focuses on learning to make predictions based on samples of target behaviors or past observations of data. Numerous machine learning algorithms are available. In this study, we utilized well-known and effective supervised machine learning algorithms, including Naïve Bayes, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Artificial Neural Networks (ANN), Linear Discriminant Analysis (LDA), and Decision Trees. The RapidMiner tool was employed for developing machine learning algorithms and conducting data mining experiments in this study.

Artificial Neural Network

An Artificial Neural Network (ANN) is a distributed and parallel information processing system modeled after human biological neural networks. It is a computational method based on the interaction of multiple fully connected processing elements (Bucak & Karlik, 2011). Typically, the Multi-layer Perceptron (MLP) architecture, as shown in Figure 2, consists of one input layer, two hidden layers, and one output layer, which are used to separate convex regions of input data. It employs the Back-Propagation (BP) training algorithm, with the tanh activation function being the most effective for all units (Karlik & Karlik, 2021). As illustrated in Figure 2, the presence of hidden units enables these networks to solve the computational complexity of associations between patterns. In our application, the optimal coefficients for the ANN (learning rate and momentum) were found to be 0.3 and 0.2, respectively, after 500 iterations. The output layer consists of four neurons labeled as pre-fatigue successful shot, pre-fatigue unsuccessful shot, post-fatigue successful shot, and post-fatigue unsuccessful shot. The input layer contains the kinematic features derived from joint angles.

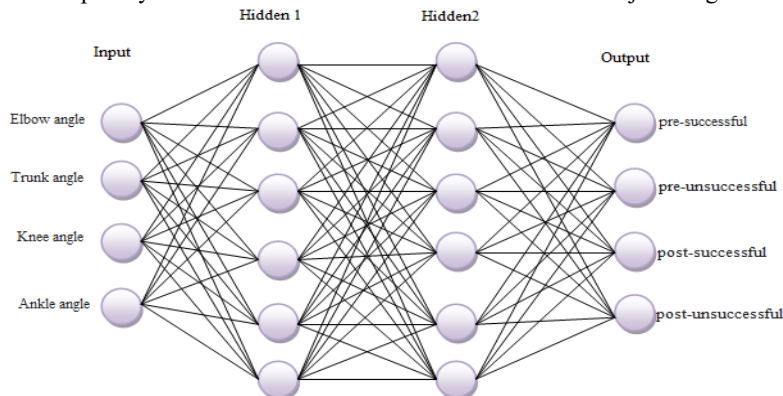


Figure 2. Used MLP architecture

Naïve Bayes

Naïve Bayes is a statistical supervised learning classifier that is simple and extremely fast compared to other more sophisticated machine learning techniques. The decoupling of the class conditional feature distributions allows each distribution to predict the value of the data independently, as a one-dimensional distribution. As a supervised learning method, the Naïve Bayes classifier is highly efficient in solving complex problems due to the precise nature of its probability-based approach (Karlik & Oztoprak, 2012). Additionally, the Naïve Bayes classifier assumes that the presence of a specific feature in a class is independent of the presence of other features. The results of the simple distribution of features for each movement in our application are shown in Figure 3.

Support Vector Machines

As a well-known statistical classification technique, Support Vector Machines (SVM) is associated with a supervised learning algorithm. SVM performs the task of finding an optimal separating hyperplane that maximizes the margin of separation between two classes. SVM analyzes data to recognize patterns used for classification and regression analysis. It provides a robust solution for multiclass art classification problems that are not linearly separable. In this study, the SVM model used consists of 53 support vectors for classifying the dataset. The parameters used are defined in Table 4.

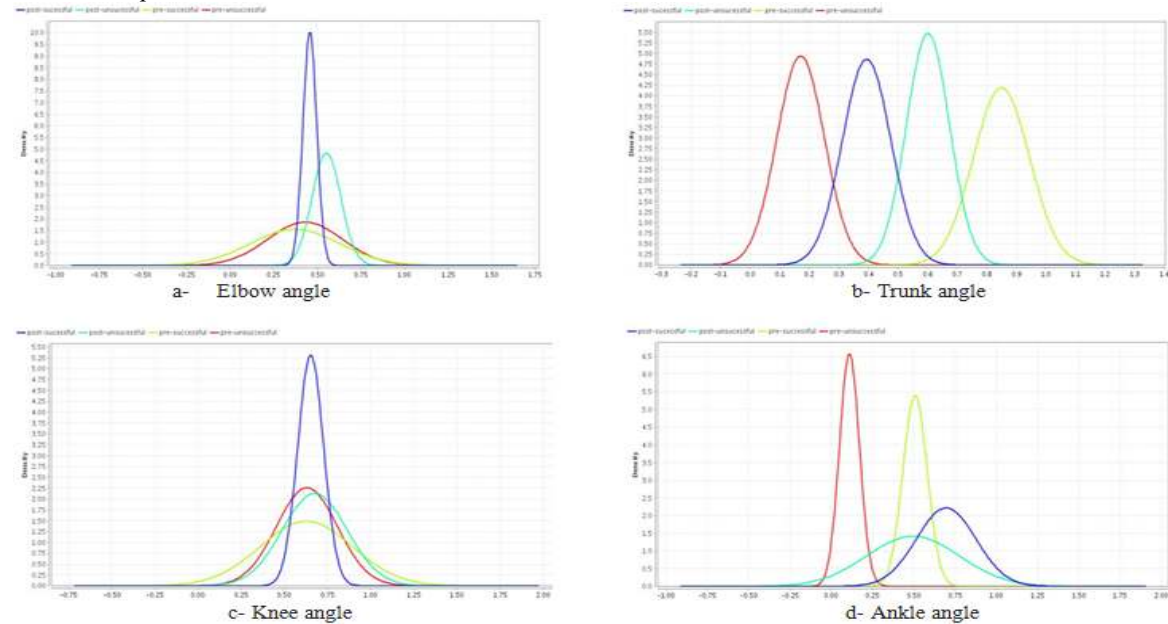


Figure 3. Simple distribution of features elbow, trunk, knee and ankle angles where (red line: pre-unsuccessful shots, yellow line: post-unsuccessful shots, green line: post-successful shots, blue line: post-successful shots).

Table 4. Parameters used with SVM

Parameters	Values
Kernel Type	Linear
Epsilon	0,001
Gamma	0

K-Nearest Neighbors

K-Nearest Neighbors (K-NN) is a statistical supervised learning classifier that uses the training dataset for classification by comparing each member of a "target" dataset. The data structure includes a classification variable of interest and several predictor variables. As a non-parametric classification method, K-NN calculates class memberships based on the k-nearest training examples. The main advantage of K-NN is that it requires fewer iterations for both training and testing. Moreover, it is an effective algorithm for problems with a low number of classes (Karlik & Hayta, 2014). In our study, the K-NN classification refers to the k value that represents the closest data points. We achieved the best accuracy when using one nearest neighbor.

Decision Tree

The Decision Tree Classifier is used in many diverse areas and is one of the possible approaches to multistage decision-making. The most important feature of a Decision Tree is its ability to break down a complex decision-making process into a series of simpler decisions, thus providing a solution that is often easier to interpret. Decision tree models are particularly useful in the field of data mining because they achieve acceptable accuracy and are relatively inexpensive to compute (Mienye & Jere, 2024). In this study, the minimum split size is 4, and the leaf size is 2. Figure 4 shows the structure of the Decision Tree used in this study based on our dataset.

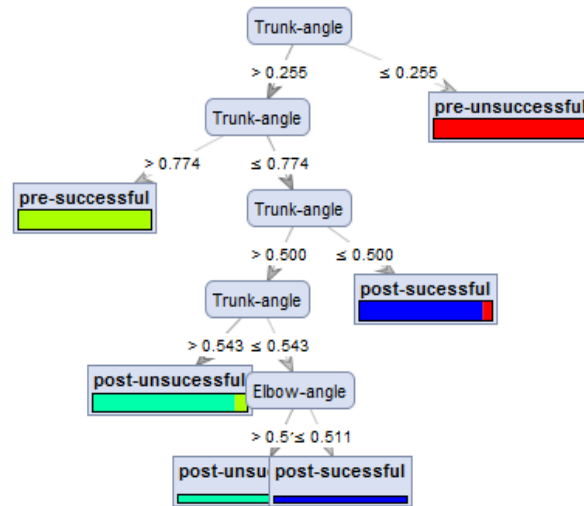


Figure 4. Structure of used Decision tree

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a well-known and useful algorithm for both data classification and dimensionality reduction. LDA effectively handles cases where within-class frequencies are unequal, and its performance is analyzed using randomly generated test data (Yilmaz et al., 2019). However, the lowest classification accuracy was obtained using this method for both datasets.

Discussion and Conclusion

In basketball, free throws, especially in the last minutes of the game, are very important as they play a role in determining whether the team wins or loses. In this context, all players must have good free throw concentration and, more importantly, good shooting mechanics. To be successful in free throws in basketball, the ball must have the appropriate kinematic characteristics with the angle of its exit from the hand, the appropriate position in the air, and optimum speed (Olteanu et al, 2023). In this study, we presented a classification of optimum Free Throw Shooting before and after fatigue using different AI (machine learning) techniques. We then evaluated the effects of principal component analysis (PCA) on classification accuracies. For this purpose, Artificial Neural Networks, Naïve Bayes, Support Vector Machines, K-Nearest Neighbors, Decision Trees, and Linear Discriminant Analysis were used as supervised machine learning classifiers. Before classification, Min-Max normalization was applied to preserve the relationships among the original data values. A problem may occur if the value of an unseen data point falls outside the $[0, 1]$ range. PCA was then used to reduce the dimensions of the dataset with minimal loss of information.

As shown in Table 5, classification accuracies were obtained using 10-fold cross-validation to evaluate the accuracy of all the machine learning techniques. Additionally, the effects of feature reduction using PCA on classification accuracies were compared with normal data, as seen in Table V. Based on cumulative variance, three principal components cover more than 98% of the total variability in the continuous data space. Finally, the performance results for each machine learning technique were compared. Naïve Bayes achieved the best classification accuracy (98.44%) using raw data. However, ANN showed the best classification accuracy (95.31%) when data reduction with PCA was used. We obtained the same accuracy for both datasets using SVM. These findings emphasize the importance of selecting the right machine learning technique for sports performance analysis, particularly when considering the complexity of player fatigue, which significantly affects shooting performance. This work contributes to the growing body of research in sports science that utilizes AI and machine learning to evaluate and predict athletic performance under varying conditions.

Table 5. Classification accuracies of used algorithms

Classifier	Classification accuracies	
	Normal data	PCA data
ANN	92.19%	95.31%
NAÏVE BAYES	98.44%	81.25%
SVM	92.19%	92.19%
K-NN	95.31%	92.19%
DECISION TREE	90.62%	76.56%
LDA	81.25%	75%

Table 6 summarizes the error rates and class precision based on the confusion matrix using the Naïve Bayes classifier for normal data and ANN for PCA-reduced data. As described in Table 6, high accuracy values and low error rates were observed. On the other hand, the highest rates of True Positives (TP) and True Negatives (TN) indicate that fatigue affects the kinematics of free throw shooting in basketball.

Table 6. Error rates and class precision based on confusion matrix where (TP: true positive, FP: false positive, FN: false negative, TN: true negative, P: class precision, ER: error rates).

Method	Shots	TP	FP	FN	TN	P	ER
NAÏVE BAYES for normal data	Post-s	16	0	0	48	100%	0.00
	Post-u	15	0	1	48	100%	0.02
	Pre-s	16	1	0	47	94.12%	0.02
	Pre-n	16	0	0	48	100%	0.00
ANN for PCA data	Post-s	15	0	1	48	100.00%	0.02
	Post-u	16	3	0	45	84.21%	0.05
	Pre-s	14	0	2	48	100.00%	0.03
	Pre-n	16	0	0	48	100.00%	0.00

The practical implications of this research are substantial for basketball coaching and training, as it offers a data-driven approach to assess players' shooting mechanics and the effects of fatigue on their performance. Understanding how fatigue alters the kinematics of free throw shooting can help coaches and sports scientists develop targeted training programs that focus on optimizing joint angles and improving shooting consistency, particularly for players who are prone to fatigue, such as pivots.

Furthermore, the theoretical contribution of this study extends the application of machine learning to sports biomechanics, presenting a novel approach for analyzing kinematic data in basketball. Future research could focus on refining the models by incorporating additional features, such as player posture or the influence of psychological factors on shooting performance for multiple players. Additionally, exploring the specific joint angles that are most critical for successful free throws could lead to the development of personalized training protocols.

In conclusion, this study not only advances the use of AI in sports performance analysis but also provides practical insights into how fatigue affects basketball players' shooting mechanics. The findings have significant implications for coaching, player development, and future research, offering a foundation for more precise, data-driven decisions in sports training and performance enhancement.

Acknowledgements

The authors would like to thank Prof. Ayhan Goktepe, The College of Technical Sciences, Selcuk University, Konya, Turkey, for generously sharing his data with us.

Conflict of Interest We confirm that there are no conflicts of interest.

References

- Abdelkrim, N.B., El Fazaa, S. & Ati, J. (2007). Time-motion analysis and physiological data of elite under-19-year-old basketball players during competition. *British journal of sports medicine*, 41(2), 69-75.
- Apriantono, T., Nunome, H., Ikegami, Y., Sano, S. (2006). The effect of muscle fatigue on instep kicking kinetics and kinematics in association football. *Journal of Sports Sciences*, 24(9), 951-960.
- Begg, R. & J. Kamruzzaman, A. (2005). machine learning approach for automated recognition of movement patterns using basic, kinetic and kin ematic gait data. *Journal of biomechanics*, 38(3), 401-408.
- Bourdas, D.I., Travlos, A.K., Souglis, A., Gofas, D.C., Stavropoulos, D., & Bakirtzoglou, P. (2024). Basketball fatigue impact on kinematic parameters and 3-point shooting accuracy: cardiorespiratory fitness associations of high-level players. *Sports*, 12(3), 63.
- Bucak, I. O., & Karlik, B. (2011). Detection of drinking water quality using CMAC based artificial neural Networks. *Ekoloji*, 20(78), 75-81.
- Chappell, J.D., Daniel C Herman, D.C., Knight, B.S., Kirkendall, D.T., Garrett, W.E., Yu, B. (2005). Effect of fatigue on knee kinetics and kinematics in stop-jump tasks. *The American journal of sports medicine*, 33(7), 1022-1029.
- Forestier, N. & Nougier, V. (1998). The effects of muscular fatigue on the coordination of a multijoint movement in human *Neuroscience letters*, 1998. 252(3): p. 187-190.
- Fukuchi, R.K., Eskofier, B.M., Duarte, M., Ferber, R. (2011). Support vector machines for detecting age-related changes in running kinematics. *Journal of Biomechanics*, 44(3), 540-542.
- Ge, C. (2024). Optimization study of a dynamic assessment model of physical fitness for youth basketball training. *Applied Mathematics and Nonlinear Sciences*. 9(1), 1-15.
- Göktepe, A., Ak, E., Söğüt, M., Karabörk, H., & Korkusuz, F. (2009). Joint angles during successful and unsuccessful tennis serves kinematics of tennis serve. *Joint Diseases and Related Surgery*, 20(3), 156-60.

- Karlık, B., (2011). Hepatitis disease diagnosis using backpropagation and the Naïve Bayes classifiers. *Journal of Infor. Technology and Engineering Sciences*, 1(1),49-62.
- Karlık, B. & Hayta Ş. (2014). Comparison machine learning algorithms for recognition of epileptic seizures in EEG. *IWBPIO-2014, Granada 7-9 April, 2014*, 1-12.
- Karlık, M. & Karlık, B. (2021). Second language learning with affective factors and deep neural networks methods. *Canadian Journal of Language and Literature Studies*, 1(3), 26–43.
- Karlık, B. & Öztoprak, E. (2012). Personalized cancer treatment by using Naïve Bayes classifier. *Inter. Journal of Machine Learning and Computing*, 2(3), 339-344.
- Karlık, B., Yilmaz, M. F., Ozdemir, M., Yavuz, C. T., & Danisman, Y. (2021). A hybrid machine learning model to study UV-vis spectra of gold nanospheres. *Plasmonics*, 16(1), 147-155.
- Kelmendi, D. S., Miftari, F., Tekin M. (2021). Kinematic analysis of the basketball free throw in the preparation phase of elite athletes. *International Journal of Human Movement and Sports Sciences*, 9(6),1204-1212.
- Li, F., Li, Z., Borović, I., Rupčić, T., & Knjaz, D. (2021). Does fatigue affect the kinematics of shooting in female basketball? *International Journal of Performance Analysis in Sport*, 21(5), 754-766.
- Mienye, I. D. & Jere, N. (2024). A survey of decision trees: concepts, algorithms, and applications. *IEEE Access*, vol. 12, 86716-86727.
- Olteanu, M., Oancea, B. M., & Badau, D. (2023). Improving effectiveness of basketball free throws through the implementation of technologies in the technical training process. *Applied Sciences*, 13(4), 2650.
- Rusdiana, A., Ray, H. R. D., Syahid, A. M., & Putra, Y. S. D. (2020). The effect of fatigue on free throw kinematic movement in basketball. In *4th International Conference on Sport Science, Health, and Physical Education*, 389-392, Atlantis Press.
- Schmidt, A., (2012). Movement pattern recognition in basketball free-throw shooting. *Human movement science*, 31(2), 360-382.
- Uygur, M., Goktepe, A., Ak, E., Karabörk, H., & Korkusuz, F. (2010). The effect of fatigue on the kinematics of free throw shooting in basketball. *Journal of Human Kinetics*, 24, 51-56.
- Yilmaz, M. F., Danisman, Y., Ozdemir, M., Karlık, B., Larour, J. (2019). Investigation of electron beam effects on L-shell Mo plasma produced by a compact LC generator using pattern recognition. *Matter Radiat. Extremes*, 4(2), 027401.
- Zhang, A. Chen, K. (2024). Simulation of image feature extraction based on optical sensors in basketball target recognition system. *Optical and Quantum Electronics*, 10.1007/s11082-023-05993-1, 56, 3.