Journal of Applied Artificial Intelligence
https://jaai.sabapub.com
ISSN: 2709-5908

*2023 Volume 4, Issue 2: 65-85*

*DOI:* https://doi.org/10.48185/jaai.v4i2.814

# Artificial intelligence, sports, and athlete's performance

**Louizi Imen[1], Rekik Malek[2], Azaiez Fairouz[3]**

[1] ISSEPS, University of Sfax, Sfax 3022, Tunisia, imenlouizi2016@gmail.com
[2] Research Unit (NTS' COM), ENET'COM, University of Sfax, Sfax, Tunisia, 22.malek@gmail.com
[3] ISSEPG, University of Sfax, Sfax, Tunisia, fairouz.kyranis@yahoo.com

Received: 07.08.2023 • Accepted: 25.11.2023 • Published: 29.12.2023 • Final Version: 29.12.2023

**Abstract:** The artificial intelligence (AI) has big changes in the sports sector in extenuating the problem of health in sports. According to studied approaches, AI learning methods and techniques can decrease accidents during practicing sports by predicting the health situation of players before the happen of any incident in the match. Thus, players can participate and complete their matches with all security owing to AI. In this paper, we introduce our approach that is predicting sport type adequate for athletes joining sportive organizations from their blood characteristics using PCA method. We do analyze and standardization of the dataset of players and then utilizing three types of classifiers in the generalization of dataset and do a comparison between them to conclude test training scores of the data. We utilized three classifiers (KNN, SVM and MLP) to do comparison between them in the evaluation of the approach results and we concluded that the best results in interpretation of feminine athletes was use of KNN algorithm and the best results of interpretation in male athletes was with MLP algorithm.

**Keywords:** Artificial intelligence, sports, athletes 'performance, AI learning methods and techniques, PCA method.

## 1. Introduction:

Accidents in sports are very frequent like heart attack, muscle sprain, etc. For instance, in the U.S., about 30 million children and teens are practicing sports, and more than 3.5 million injuries each year are measured by the participants, which cause some loss of time for participation and performance of players [1].

According to many studies, the use of artificial intelligence, like deep learning methods, for prediction of sporting results (outcomes) shows how the behavior of teams can be modeled in different sporting contests using multi-layer ANNs (Artificial Neural Networks).

In addition, authors in a study show how algorithms of AI can be used to develop an artificial trainer to recommend athletes with an informed training strategy after taking into consideration various factors related to the athlete's physique and readiness.

According to an existing academic work [2], in a range of sports, authors have summarized tackling issues such as match outcome modeling, in-game tactical decision making, player performance in fantasy sport games, and managing professional players' sport injuries.

According to studies on immunity of athletes, authors have concluded that intensive and prolonged physical effort lead to impaired immunity up to 72 hours after the exercise.

Furthermore, researchers assumed that the psychological factor of players when preparing to big events has a role in the relationship between immunity and sports. According to Ashley Ying-Ying Wong and ALL [3], they have utilized a small example of video footage of professional football players that were analyzed to track each player's time of close body contact and frequency of infection-risky behaviors to investigate the risk of virus transmission during football games.

In this paper, we focus on the inventions of AI in the sports sector concerning performance of athletes in matches. We outline several articles by showing the existing solutions of AI, the implementation phase utilized in each approach and its limits by specifying advantages and disadvantages. We also take a comparison between the analyzed approaches and that by specifying the employed device, the shortage of the solution and the results of the implementation.

Finally, we propose our approach that is about predicting the type of sport adequate for each athlete by introducing their blood variables and physical characteristics and employing PCA method in the clustering of data. The approach aims to facilitate the way to people who want to practice sport according to their characteristics and physical properties and simplify big time to be expert in a type of sport.

## 2. Related works of artificial intelligence in sports:

According to readings in current articles, many approaches of artificial intelligence have been developed to deal with sport and the problem of player's performance. In each approach, we'll distinguish the method of machine learning utilized in the implementation phase, the database and the employed device. Also, we'll mention the advantages extracted from each study.

### 2.1 Artificial Intelligence for Sport Activity Recognition, 2019:

Sergei Bezobrazov and ALL [4] have employed the PIQ ROBOT device that is a smart watch fixed on the wrist of the tennis player during the game to detect forehands, backhands and serves of the player and transferred sensor data to the PIQ log application installed in a smart device via BLE (Bluetooth Low Energy).

The artificial neural network was the area of research of the implementation in the author's approach. The authors aim to reduce the time complexity of the approach by using the sliding window algorithm in the start of the execution approach (in the first layer of neural network); this technique shows how a nested for loop in some problems can be converted to a single for loop. In addition, they summarize large datasets to a smaller number of codebook vectors suitable for classification in their approach by introducing the hidden layer training algorithm that utilizes the Euclidian distance for input and weight vectors of neurons.

They also demonstrate the method of LVQ-nets (Learning Vector Quantization) in execution of artificial neural network in the hidden layer to solve the problem of tennis motions classification and segmentation. The LVQ is a prototype based supervised classification algorithm and it applies the winner take-all based approach. In winner-take-all training algorithms, one determines, for each data point, the prototype which is closest to the input according to a given distance measure. The position of this winner prototype is then adapted, for instance the winner is moved closer if it correctly classifies the data point or moved away if it classifies the data point incorrectly.

The competitive learning rule (winner-takes-all) is used for training of hidden layer. The number of neurons in hidden layers equals to the number of existing classes. The acquisition of the data in this approach was divided into two systems: vision-based and sensor-based.

The authors distinguished 9 classes of different strikes in tennis that are grouped into 4 types. For the training process of neural network, we'll introduce parameters, and every motion (movement) has a different period in the device. Concerning the description of system structure, the authors employed 4 neural networks. In the test of trained neural network, the authors have utilized control files for two players. The accuracy for player 1 was better that the player because in the first the data was used in the training and in the second player the data is new for classifier.

*Advantages:*

The model learning is trained significantly faster than other neural network techniques like Back propagation. In fact, it can approximate any classification problem and the attributes can be compared using a meaningful distance measure.

*Limits:*

- The approach didn't solve the problem of tennis motions classification when using close approximations in measurements of some motions like flat and lifted classes.
- In case of limited power resources, the implementation is very critical.
- The problem with vision-based systems in database is the high cost of the GPUs, the range of use is limited by the camera volume and the big number of cameras used.

**2.2 A Systematic Literature Review of Intelligent Data Analysis Methods for Smart Sport Training, 2020:**

Alen Rajšp and ALL [5] were interested in the domain of smart sport training (SST) that became necessary in the future direction of research in sport. The SST is a type of sport training where a coach is replaced by a smart agent like sensors, wearables and internet of things devices (loT) or intelligent data analysis (extraction of useful knowledge from data) that improve the training performance or reduce the workload when maintaining the same or better training performance. The authors have reviewed the latest advances in the development and use of intelligent data analysis methods in the domain of SST.

This process allows the athlete to choose from a variety of possible training regimes without the need for employing an actual person to aid them in their training. The objective of this review is to identify how modern smart applications and methods assist athletes and trainers in sports training and how fast the theoretical knowledge is transitioning into practical real-world use cases (determine which sports are the most supported by these methods).

Also, the authors have divided the studies of the review by sport type. Studies show that the SST is and can be a determining force in transforming all the four training phases. The used datasets were private and from real data.

According to the results found, the most used intelligent data analysis methods from computational intelligence algorithms keep support vector machines, artificial neural networks, k-nearest neighbors, and random forest.

The authors outline the challenges faced in the future and they are:

- Knowledge transfer into the real-world and validation level research.
- Cooperation with trainers and athletes.
- Obtaining test datasets and their propagation.

*Advantages:*

The research is addressed to various destinations in sports sector such as: sport training, sport trainers, or sport trainees.

Furthermore, it addressed sport as an athletic activity requiring skill or physical talents and often of a competitive nature. Also, this SLR used at least one of the intelligent data analysis technologies (e.g., data mining, computational intelligence, big data, and machine learning).

*Limits:*

- The databases were restricted on some criteria (number of consecutive pages, indexing dates, etc.).
- The research was limited to the five scientific databases/search engines: ACM Digital Library, IEEEXplore, ScienceDirect, Scopus, and Google Scholar

**2.3 Artificial intelligence in sports, 2019:**

The authors in [6] determine many tools of employ of AI in sport like the sports analytics which is a sporting tool that can provide player evaluation separately and as a team. This tool is based on analytics of big data and digital marketing.

They present the objective of AI in sports; it's to mimic the cognitive human tasks such as teaching and solving problems. They have also talked about wearable tools like the Fitbit bounding. AI is helpful to enable equipment to provide a quantified self that enables athletes to enhance their training and results. Finally, they précised an objective for the approach to analyze the stats of players and predict how is beneficial the player for the team. They took the example of the cricket game.

The location (stadium), town, toss winner and toss decision (field/bat) could be some variables that influence game results. There were five metrics to analyze the batsman: hard hitting ability, finisher, fast scoring ability, consistency and running between wickets. And five metrics for the bowlers: economy, wicket taking ability, consistency, crucial wicket taking ability and short performance index.

The MVPI (The Motives, Values, Preferences Inventory) works like that: calculate metrics (identify metrics ranges), feature selection and player ranking. According to the result's data, 50% from the time the team who won the toss also won the game, so it's not enough to determine the game. But according to the location, it's a characteristic more significant than the winner of the toss results. Thus, the model will predict the winner by the result variable.

*Advantages:*

The predictive analytics select winners and help better control risk. Moreover, insights and quantifications of information provide accurate and timely responses. Finally, the input information automates graphs, reports, and predictive models for updates.

*Limits:*

- More variables of predictors can lead to invisible training information.
- The result of invisible training information is overfitting.

- The customer must balance training set variables and predictors based on precision and cross-validation score.
- According to the result's data, 50% from the time the team who won the toss also won the game, so, it's not enough to determine the game.

### 2.4 Neuro-fuzzy analytics in athlete development (NueroFATH): a machine learning approach, 2021:

This paper [7] presents a machine learning approach to the physical assessment (evaluation) of athletes known as NueroFATH. This paper aims to utilize the assessment parameters in determining the athletes' category in groups such as gold, silver, and bronze.

The NueroFATH model employs both supervised and unsupervised learning algorithms to carry out the physical assessment. The supervised learning contains neural networks (i.e., multilayer perception) that deliver the reliable mining (feature extraction and representation) of collected data for the classification unlike other traditional supervised learning, including support vector machine (SVM) and logistic regression.

The unsupervised learning includes K-means and fuzzy C-means (Fuzzy C-means is a method of clustering which allows a data point to belong to two or more clusters) that help to recognize the pattern of athlete features and lower the potential in accuracy in labeling the dataset by performing learning over the unlabeled data.

The work presented in this paper can not only be used to identify the best athletes, but also determine which physical characteristics are important to classify them in various categories.

According to the results obtained, authors infer that there are differences between supervised (neural networks) and unsupervised learning (fuzzy c-means and K-means).

Advantages:

The work presented in this paper can not only be used to identify the best athletes, but also determine which physical characteristics are important to classify them in various categories.

It will not only save cost and time used in training and coaching the athletes, but also can be utilized to invest in the features which would eventually lead to higher quality and quantity of medals won.

*Limits:*

The framework can be enhanced by employing a semi-supervised learning technique that supports the generation of an efficient learning model by employing learning over both unlabeled and labeled data.

### 2.5 Using Artificial Intelligence for Pattern Recognition in a Sports Context, 2020:

Pattern recognition [8] is the ability of a machine to recognize patterns in their environment, using artificial intelligence learning abilities to make decisions.

The study is interesting in the classification of action in sports using wearable technology combined with a deep learning method (i.e., LSTM) and an ensemble classification method (i.e., DBMM), as well as comparing these two methods with the ANN. The authors developed architecture for data integration coming from two AI inventions: TraXports and MBody3.

The sport performed in this study was the futsal sport that is a mini football team that contains five players including the goalkeeper. The methodology of this study consists of wearing the two AI inventions (TraXports is worn by each player and MBody3 is worn by only one player). Then a data acquisition (physiological and kinematic data) is required from authors to build models and testing them to lastly classify actions of the athletes that are running, running with the ball, passing, walking, walking with the ball, shooting, and jumping.

According to the results found, the feature selection (choice of the classification method) was the base of the pattern recognition study to improve the performance of an athlete.

*Advantages:*

The feature selection is an important step for the pattern recognition study that allied to an aware choice for the classification method as well as its parameterization, the overall classification performance can be improved.

*Limits:*

- The few used trials in the study because the augmentation of trials makes contribution for LSTM and ANN methods.
- When tested alone, ANN couldn’t give the expected results but when combined with another classification method (as it is the case in DBMM) the combined solution can achieve a higher classification performance.
- A convergence analysis between methods should be realized to evaluate which of the two classifiers converge to the adequate action the fastest.

**2.6 Research on basketball players training strategy based on artificial intelligence technology, 2020:**

The research [9] of artificial intelligence technology can optimize the model of basketball players’ physiological indicators, in addition it can solve the problem between overtraining and effective training and can help basketball players develop more appropriate training strategies.

Due to this research, basketball coaches can provide the system sports assistance as following:

1. Understanding and evaluation of a pair of athlete’s physical fitness and athletic ability
2. The evaluation of basketball players sports technology
3. The ability to supervise the implementation of athletes to provide targeted training
4. Make appropriate adjustments and optimizations to the entire sports plan based on the actual situation of the basketball players.

*Advantages:*

Basketball sports technology and sports mode, build a multi-target feedback training method, and improve the level of basketball players.

The basketball coaching is more scientific, standardized, and intelligent, monitors athletes' basketball player data in real time, corrects the mistakes made by basketball players in training, and promotes the improvement of athletes' sports skills.

It promotes the fairness and science of sports, and it has great value and significance for single athletes or collective sports.

We'll interest in the two following articles of the coronavirus pandemic that is nowadays a rampant phenomenon with AI and sport. In the next two articles, we'll focus on AI and medicine sector in relation with covid-19.

*Limits:*

- The application is a huge and difficult task.
- The system makes the training effect more obvious and has immeasurable application value.

**2.7 The application of machine learning and deep learning in sport: predicting NBA players' performance and popularity, 2022:**

In [10], authors explain the effectiveness of applying ML and DL on sport domain in basketball. Authors decided to extend further in providing information about the understanding of each ML model and its mechanism.

CRISP-DM methodology is used as the reference to construct ML and DL models. The greatest benefit is to provide a common method for communication that helps to connect a variety of technical tools and people with different skills and backgrounds to progress an efficient and effective project.

Two potential solutions were proposed: under-sampling and over-sampling techniques. Moreover, the priority in this data is to detect potential 'All-Star' players, two metrics: ROC AUC and Recall scores are our primary options. Overall, some popular evaluation metrics in ML field were selected to assess and compare our ML algorithms:

(1) Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for the regression models

(2) Accuracy, Precision, especially Recall and Receiver operating characteristic area under curve (ROC AUC) and F1 scores for the classification models.

*Advantages:*

Machine learning can give many advantages in Sport domain with its capability to predict the future outcomes, which can be seen through the study's results: RMSE of 2.1969 and MAE of 1.6465 for Regression Analysis, Recall of 0.9368 and ROC AUC of 0.9152 for Classification Analysis.

*Limits:*

As a team-oriented sport, there are some other factors influencing the players' performance and popularity (team's tactical style, coach decision, team chemistry) which are not included in this study and would be more accessible to study to improve the models in the future.

**2.8 Personalized machine learning approach to injury monitoring in elite volleyball players, 2022:**

In the study [11], authors implemented a machine learning approach to investigate individual indicators of training load and wellness that may predict the emergence or development of overuse injuries in professional volleyball.

In this retrospective study, data were collected from 14 elite volleyball players (mean ± SD age: 27 ± 3 years, weight: 90.5 ± 6.3 kg, height: 1.97 ± 0.07 m) during 24 weeks of the 2018 international season. Training sessions and matches were monitored during 24 weeks of the 2018 international volleyball season for national teams.

Based on training load and wellness indicators, we identified subgroups of days with increased injury risk for each volleyball player using the machine learning technique Subgroup Discovery. We demonstrate that the emergence and development of overuse injuries can be better understood using daily monitoring, taking into account interactions between training load and wellness indicators, and by applying a personalized approach.

In the machine learning analyses, the sample size, i.e. the number of days on which a player is monitored, was 94 ± 18 data points per athlete (mean ± SD).

Authors have distinguished that daily monitoring of overuse complaints decreases the prevalence of substantial issues. Second, we have found that jump load predictors are important for overuse issues in volleyball. Daily monitoring of overuse complaints decreases the prevalence of substantial issues. Second, we have found that jump load predictors are important for overuse issues in volleyball.

With single predictors, the authors findings show that subgroups with high severity scores could be detected based on single predictors for each of the OSTRC questions and for most of the volleyball players.

In addition to single predictors, they also performed Subgroup Discovery to detect subgroups using multiple predictors.

Authors have found individual differences in the most relevant descriptor of the training load and the time window that is considered. This suggests that every athlete has his own predictor for monitoring overuse issues.

*Advantages:*

As their personalized findings are easy to interpret, these can be used to minimize injury risks when designing training schemes for each individual player.

*Limits:*

- The overuse issues occurring on a certain day might affect subsequent training activities.
- Only in 1.3% of the cases a player had to reduce his training activities to at least a moderate extent.
- No explicit distinction between illness and injuries per body part. Authors distinguished between general symptoms (complaints) and issues that affect training participation or performance, but it would be interesting to consider for example knee or shoulder issues separately.

**2.9 Sports medicine and artificial intelligence, 2021:**

In the study [12], the authors review the use of AI-based techniques in sports medicine as a preparation for what is to come as it relates to athlete injury prediction, the interpretation of medical imaging, patient-reported outcome measures (PROMs), the delivery of value-based care and the improvement of the patient experience.

For the practicing sports medicine surgeon, authors may be able to produce a suitable algorithm that decides when to use biologic agents for the treatment of musculoskeletal conditions, though they may never have the specific knowledge guiding this decision in specific patient scenarios.

Implementation of AI in orthopaedic surgery has done a great promise in predicting injury risk. Some examples of the utility of ML in orthopaedics include the identification of fractures from radiographic images, the ability to identify orthopaedic implants for the purpose of facilitating hardware removal and/or revision, the prediction of postoperative opioid use after total hip arthroplasty and the evaluation of remote monitoring systems.

Authors have invented a remote patient monitoring system that depicts the wearable knee sleeve transmitting motion data to the smartphone, which then transmits these and all other data (steps, patient-reported outcome measures, opioid use) to the dashboard.

The data are then analyzed by the machine learning algorithm and instantaneously transmitted back to the patient while being stored on the care team dashboard. This offers the ability to improve perioperative data capture despite remotely caring for patients.

*Advantages:*

The future practice of orthopaedic surgery (Orthopaedic surgery: relating to the branch of medicine dealing with the correction of deformities of bones or muscles) necessitates that surgeon's gain sufficient familiarity with AI and ML concepts, seizing the opportunity to leverage this powerful technique and take a participatory role in its responsible deployment.

## 3. Experimental and computational details

Our contribution is to propose a new authentication system for new athletes joining a sportive organization and wish to pick the kind of sport that suit with their physical properties and also their blood properties (blood cells, iron, etc.).

### 3.1 Measurements

To begin the new approach, we did the collect of data base containing the blood characteristics, the gender of athlete, his weight, height, muscular mass, etc. Our purpose is to predict the kind of sport that deals the best with the corporal nature of the athlete either masculine or feminine. The data is implicated to 202 athletes.

The number of red cells (« rcc ») in $10^{12}l^{-1}$,

Number of white blood cells (« wcc ») in $10^{12}l^{-1}$,

Hematocrit: percent of red blood cells volume in blood (« hc »),

The concentration of hemoglobin is an existent protein in red blood cells and kept oxygen (« hg ») in deciliter per gram,

The ferritin: Size of iron stock in blood (« Ferr » in ng $dl^{-1}$,

« BMI » or « IMC » in French: Index of corporal mass : measure of weight in rapport with size kg/m² (corporal fat),

« ssf » (sum of skin folds in the skin),

« pcBfat » (percent of fat in the body),

« lbm » (thin corporal mass) in kg,

Height (« ht »), weight (« wt »), gender (« sex »)

Kind of sport (« sport »):

The available types of sports in the data base are 10 types : baseball, row (« Row »), swim (« Swim »), field « Field » (bike race, shooting, archery, etc.), 400 meters running « T-400m », free running « T-sprnt », tennis, gymnastic « Gym », running cars Polo « W-polo », basketball (« Netball »).

These data are collected in term of a study of diverse characteristics of blood varying by sport, size and gender of the athlete. Given that the theme of this data is the science and medicine in sport and exercise.

### 3.2 Non supervised learning method PCA

The ACP method allows calculating matrix to project variables in a new space using a new matrix that shows the degree of similarity between variables.

The contributions of each individual to the inertia of their cloud allow detecting the most influent observations. The contributions allow identifying influencing individuals that could determine the orientation of certain verified axes, characterized then eventually considered as additional in another analyze.

### 3.3 Analyse of blood variables

The environment associated to the implementation of our approach is the navigator « anaconda » JUPYTER notebook. It's about a notebook data processing and interactive environment that modify and execute readable documents by human by describing the data analyzing. The programmation language employed by « anaconda » is Python.

Thus, we propose to analyze blood variables by gender o athlete and kind of sport.

We take the example of the number of red blood cells (« rcc »), the average, the variance, max and min, median and std (calculation function of standard deviation of a set of data, columns, etc. to construct the curve of analyzed variable « rcc » compared with feminine gender, masculine and type of sport as indicated in the tables 1, 2 and 3 bellows.

**Table 1.** "Measures of construction of the curve rcc/feminine"

| | max | min | moyenne | variance | mediane | std |
|---|---|---|---|---|---|---|
| rcc | 5.33 | 3.8 | 4.4045 | 0.102974 | 4.385 | 0.320896 |

**Table 2.** "Measures of construction of the curve rcc/masculine"

| | max | min | moyenne | variance | mediane | std |
|---|---|---|---|---|---|---|
| rcc | 6.72 | 4.13 | 5.026569 | 0.123007 | 5.015 | 0.350723 |

**Table 3.** "Measures of construction of the curve rcc/sport"

| | max | min | moyenne | variance | mediane | std |
|---|---|---|---|---|---|---|
| B_Ball | 5.24 | 3.96 | 4.570000 | 0.130283 | 4.510 | 0.360948 |
| Row | 5.40 | 3.91 | 4.677838 | 0.148190 | 4.630 | 0.384954 |
| Netball | 4.83 | 3.80 | 4.227391 | 0.059075 | 4.230 | 0.243053 |
| Swim | 5.34 | 4.07 | 4.714545 | 0.186350 | 4.780 | 0.431683 |
| Field | 5.93 | 4.48 | 5.024211 | 0.128426 | 5.010 | 0.358365 |
| T_400m | 5.49 | 3.90 | 4.691034 | 0.177124 | 4.710 | 0.420861 |
| T_Sprnt | 6.72 | 4.64 | 5.193333 | 0.270510 | 5.000 | 0.520105 |
| Tennis | 5.66 | 4.00 | 4.790909 | 0.283329 | 4.970 | 0.532287 |
| Gym | 4.53 | 4.09 | 4.307500 | 0.041092 | 4.305 | 0.202711 |
| W_Polo | 5.34 | 4.63 | 5.032353 | 0.045069 | 5.030 | 0.212295 |

## 4. RESULTS AND DISCUSSION

The measures gave results as indicated in the following curves.

**5.1 Results of rcc analysis**

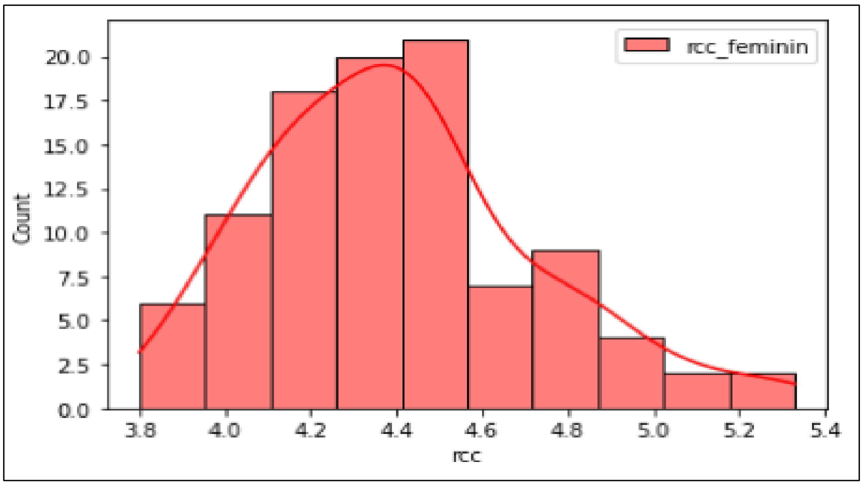


**Figure 1**. "Analysis of the rcc variables compared with feminine gender"

As seen in Figure 1, among feminine athletes, the number of red blood cells varied between 4 and 4.6 in $10^{12}l^{-1}$ (the number of blood cells per liter is $10^{12}$).

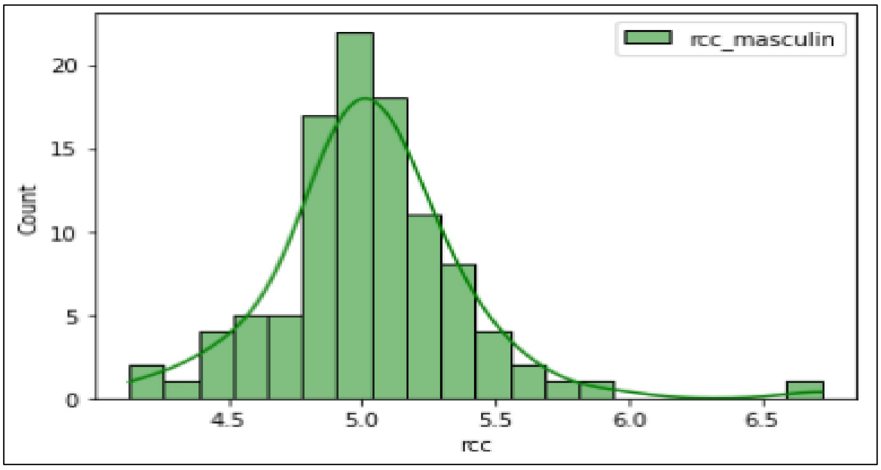


**Figure 2.** "Analysis of rcc variable compared with masculine gender"

In Figure 2, among masculine athletes, the number of red blood cells varied between 4.8 and 5.4 in $10^{12}l^{-1}$. We observe that for masculine athletes, the value of rcc variable is much elevated than feminine athletes. Then, we calculated the kernel density estimation (KDE) that is an algorithm calculating a probability model of distribution generalizing data base.

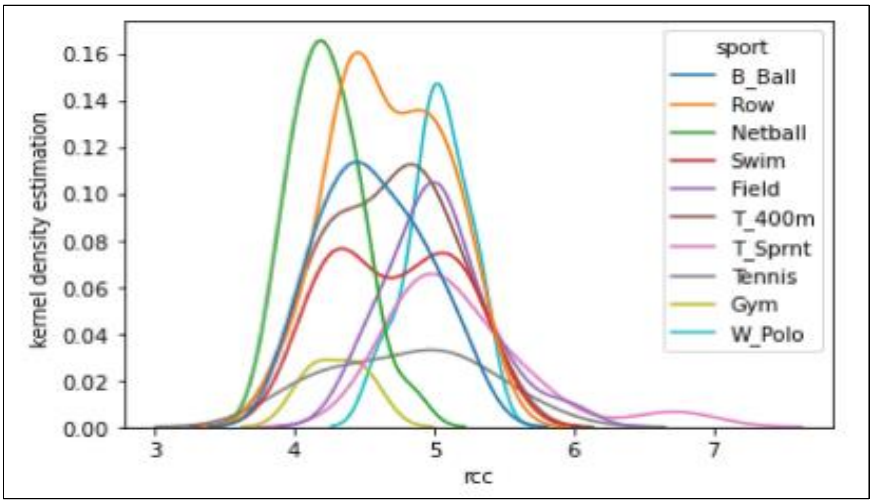


**Figure 3.** "Analysis of rcc variable compared with sport variable"

In Figure 3, according to this curve, there are three types of sports that need more number of red blood cells in the blood like "W_Polo", "Row" and "Baseball". Other sports as « Baseball » and « Field » have less red blood cells in the blood.

In the following section, we pass to the grouping and classification phase of variables to reduce the most possible the number of variables in the database employing the ACP method.

### 5.2 Results with PCA method

We begin by the decomposition of data as attached in Table 4 below.

**Table 4.** "Measures of decomposition of data"

| | rcc | wcc | hc | hg | ferr | bmi | ssf | pcBfat | lbm | ht | wt |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3.96 | 7.5 | 37.5 | 12.3 | 60 | 20.56 | 109.1 | 19.75 | 63.32 | 195.9 | 78.9 |
| 1 | 4.41 | 8.3 | 38.2 | 12.7 | 68 | 20.67 | 102.8 | 21.30 | 58.55 | 189.7 | 74.4 |
| 2 | 4.14 | 5.0 | 36.4 | 11.6 | 21 | 21.86 | 104.6 | 19.88 | 55.36 | 177.8 | 69.1 |
| 3 | 4.11 | 5.3 | 37.3 | 12.6 | 69 | 21.88 | 126.4 | 23.66 | 57.18 | 185.0 | 74.9 |
| 4 | 4.45 | 6.8 | 41.5 | 14.0 | 29 | 18.96 | 80.3 | 17.64 | 53.20 | 184.6 | 64.6 |

The table 5 explains the normalization and standardization of data to reduce variables that possess approximated values in interval [-1, 1].

**Table 5.** "Standardization of data"

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | -1.392161 | 0.299952 | -1.141877 | -1.371511 | 0.098699 | -0.544085 | 0.656960 | 0.350360 | 1.223246 | 2.598011 | 1.064207 |
| 1 | 0.017226 | 0.774183 | -0.873831 | -0.936111 | 0.358434 | -0.502209 | 0.469910 | 0.636045 | 0.530686 | 1.841996 | 0.649868 |
| 2 | -0.828406 | -1.182022 | -1.563093 | -2.133462 | -1.167509 | -0.049185 | 0.523353 | 0.374321 | 0.067528 | 0.390933 | 0.161869 |
| 3 | -0.922365 | -1.004186 | -1.218462 | -1.044961 | 0.390901 | -0.041572 | 1.170605 | 1.071024 | 0.331775 | 1.268887 | 0.695906 |
| 4 | 0.142505 | -0.115001 | 0.389816 | 0.478940 | -0.907774 | -1.153193 | -0.198124 | -0.038540 | -0.246084 | 1.220112 | -0.252471 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95 | 0.800218 | -0.352117 | 0.160062 | 0.370090 | 1.689576 | -1.377801 | -1.448091 | -1.732375 | -1.882383 | -2.035633 | -1.983488 |
| 96 | -0.671808 | 1.189136 | -0.567492 | -0.174160 | 0.390901 | -1.164613 | -1.290732 | -1.417200 | -1.734288 | -1.913695 | -1.799337 |
| 97 | 0.393062 | -1.182022 | 0.083477 | 0.478940 | -0.518171 | -1.598602 | -0.895849 | -0.976692 | -2.409425 | -2.157571 | -2.167639 |
| 98 | -0.985005 | -1.241301 | -1.716262 | -1.153811 | 0.293501 | -1.880315 | -1.347144 | -1.612572 | -2.981476 | -3.133075 | -2.720091 |
| 99 | 0.048546 | -0.352117 | 0.887616 | 1.023191 | 0.196100 | -0.639258 | -0.833499 | -0.808968 | -2.303436 | -3.120882 | -2.047941 |

100 rows × 11 columns

Moreover, we have to calculate the variance and co-variance of principal components obtained by ACP method.

Two principal components PCA1 and PCA2 have been obtained. The following curve reduces the number of BD variables and posts the kind of sport compared with the two components.

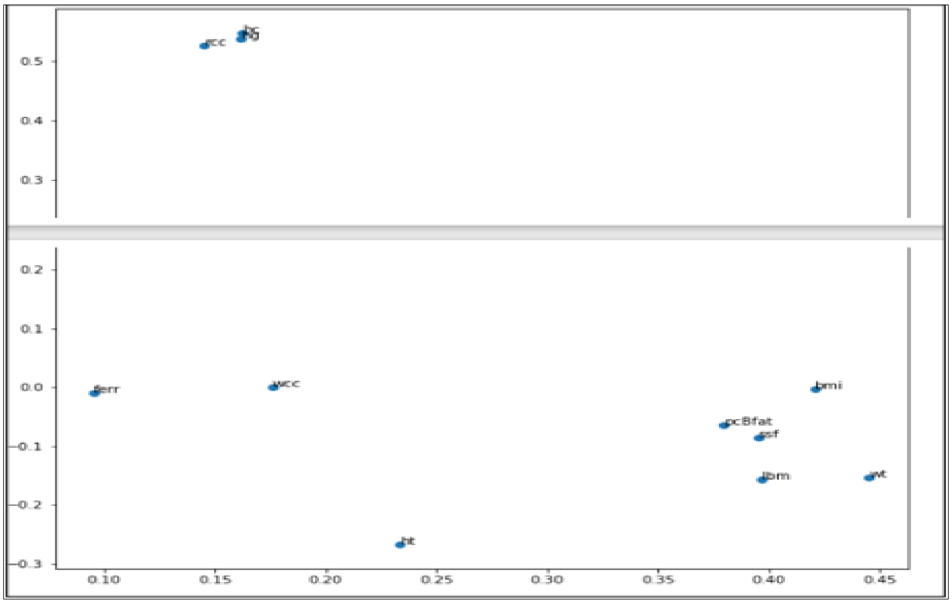


**Figure 4.** “Analysis of blood variables by ACP method for masculine athletes”

According to Figure 4, we observe that the variables « Ferr » and « wcc » (the size of the iron stocks and the number of white blood cells) in the left side of the figure are the determinants of the same type of masculine athlete according to the analysis of ACP method. When the variables "pcBfat","bmi" and "ssf” in the other side of the figure, specify another type of masculine athletes different from the first type practicing another sport.

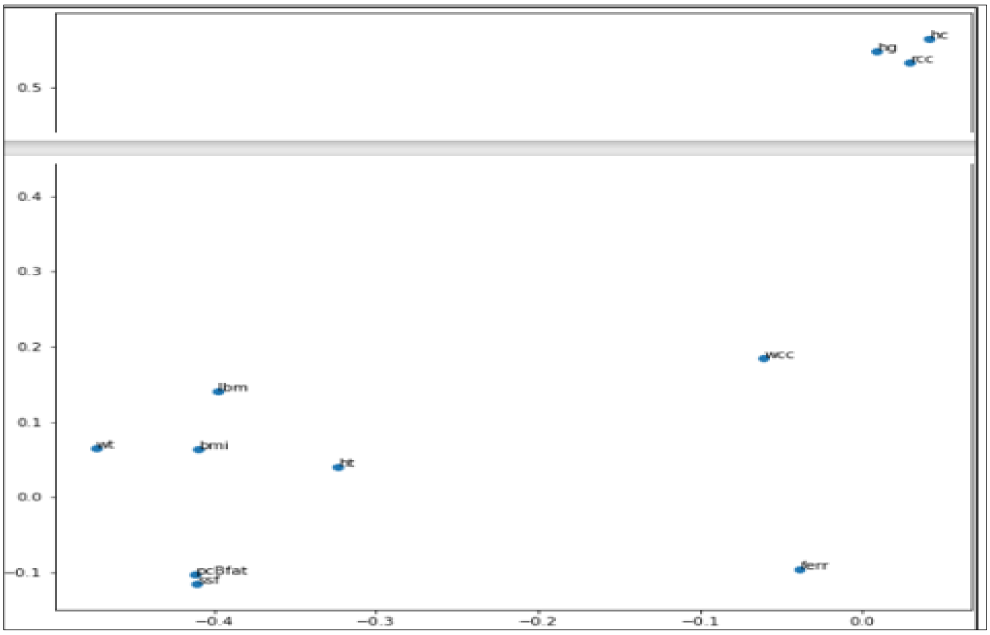


**Figure 5**. “Analysis of blood variables employing ACP method for feminine athletes”

According to Figure 5, we distinguish that the variables « lbm » (thin corporal mass), « bmi » (corporal fat) and height « ht » in the other side of figure are the determinants of same type of feminine athletes depending of the analysis of ACP method. And the variable « wcc » (number of white blood cells) of the other side in the figure specify another type of feminine athletes different from the first type practicing another kind of sport.

The following figures explain curves with reference to the analysis of variables with ACP method; and that show the blood variables are affecting on athletes (masculine and feminine) on their choices of kind of sport.

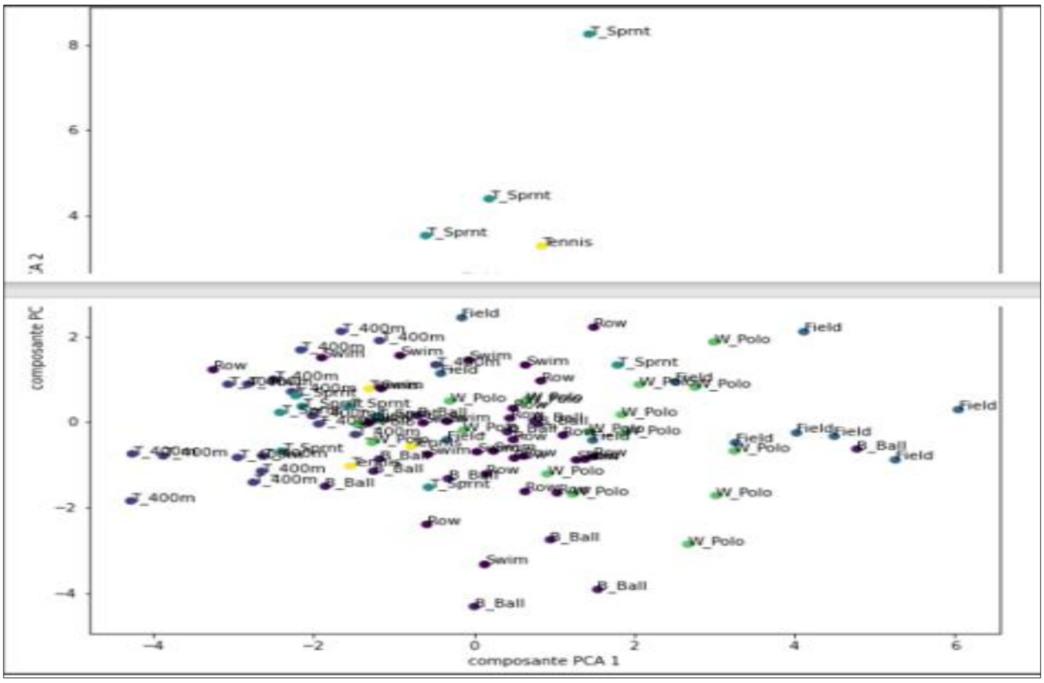


**Figure 6.** “Analysis of sport type with ACP method for masculine athletes”

In figure 6, we distinguish that on the left side the kind of sport called « T_400m » (running of 400 meters) and « T_sprnt » (free running) definite a same type of masculine athletes that can practice that kind of sport according to their corporal characteristics. On the other side, we observe that « Field » and « W_Polo » specify the same type of masculine athletes depending on blood characteristics.

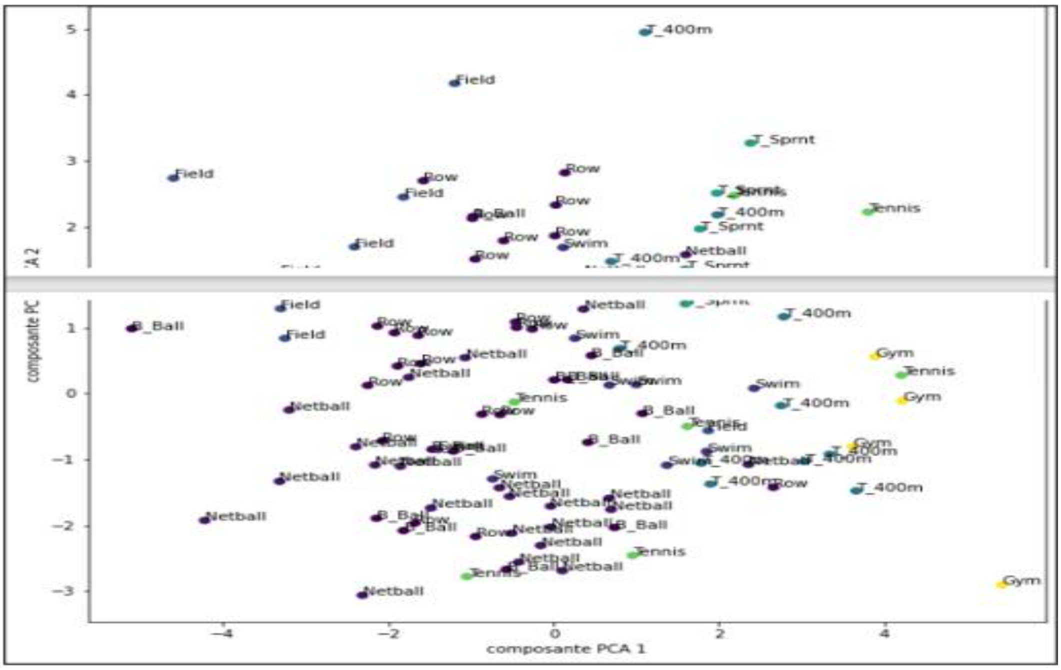


**Figure 7.** "Analysis of sport type by ACP method for feminine athletes"

As seen in Figure 7, we notice that in the left side of the figure the type of sport called « Netball » (basketball)and « Field » (open door sport) specify a same kind of feminine athletes that allow practicing that type of sport according to their corporal characteristics. In the other side, we observe that « T_400m » and « Gym » specify the same type of feminine athletes according to their blood characteristics.

### 5.3 Interpretations

We conclude that in masculine athletes, running need more stocks of ferritin (ferritin plasma) and the number of white blood cells to avoid anemia and increase immunity in masculine athletes.

As well the open doors sports (bike race, shooting, archery, etc.) and the « Polo » cars races need an elevated percent of fat in the athlete body, sum of skin folds and corporal fat.

While in feminine athletes, outdoor sports and basketball need elevated thin corporal mass (amount of water between 70% and 73%), corporal fat (weight have a big importance in the performance) and the height (especially in basketball players).

Moreover, running and gymnastic need elevated number of white blood cells to increase players' immunity.

***Methods***

We utilized three classifiers in our approach that are: KNN, SVM and MLP.

- *KNN classifier:*

The evaluation phase of the KNN classifier explains the display of the evaluation scores and the graphical representation of the scores (training score and test score) in relation to the values of the k neighbors. The purpose of this evaluation is to observe the result of the generalization of the model.

The KNN classification is made for sports prediction for both female and male athletes. The comparison between the generalization models of male and female athletes.

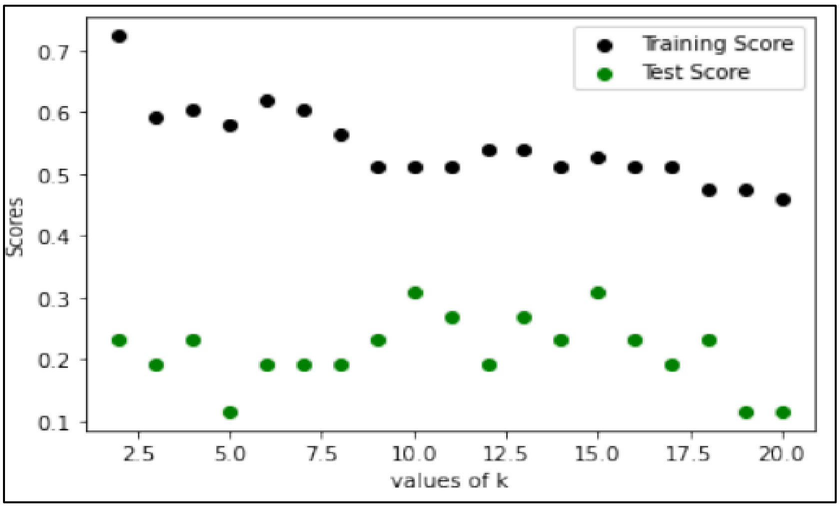


**Figure 8.** "The generalization of the KNN model for male prediction"

In Figure 8, the k values in training (the data trained in the model) score is higher than in test score (the actual scenario of the model). We conclude that the generalization is weak.

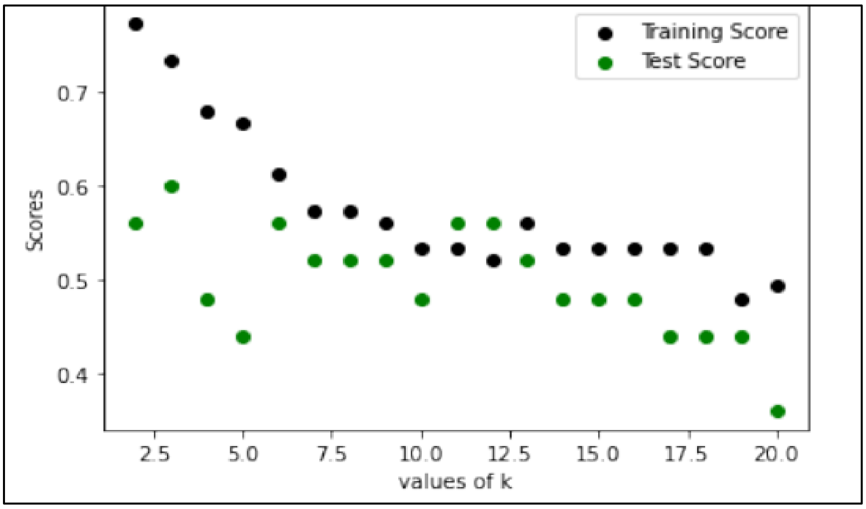


**Figure 9**. "The generalization of the KNN model for female prediction"

In Figure 9, the values of k in training score are proportional to those in test score. Hence the generalization is acceptable. It can be seen from the two figures 8 and 9 that the evaluation of classifying KNN for female athletes is much more beneficial than for male athletes.

- *SVM classifier:*

Its parameters are: the regularization parameter C, the kernel ("kernel"): the types of the kernel are linear, sigmoid and poly. In the implementation, we have displayed the evaluation scores for the linear, poly and sigmoid activation functions. The graphical representation of the scores for the three activation functions is shown below for female athletes and for male athletes.

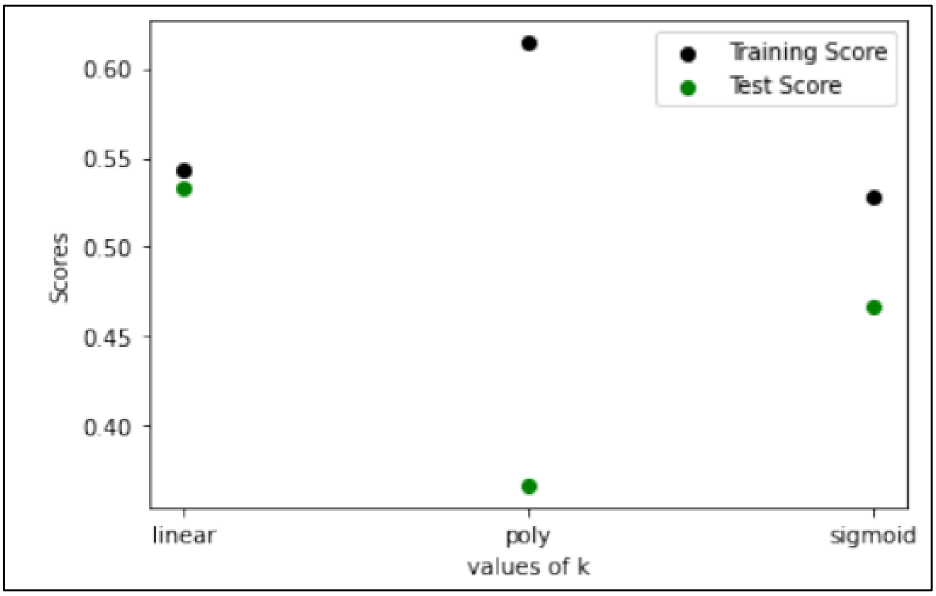


**Figure 10.** “SVM classification scores for female athletes”

In Figure 10, the training score for the three functions is higher than in the test score. We conclude that the generalization of the model is weak for the poly and sigmoid functions and strong for the linear function.

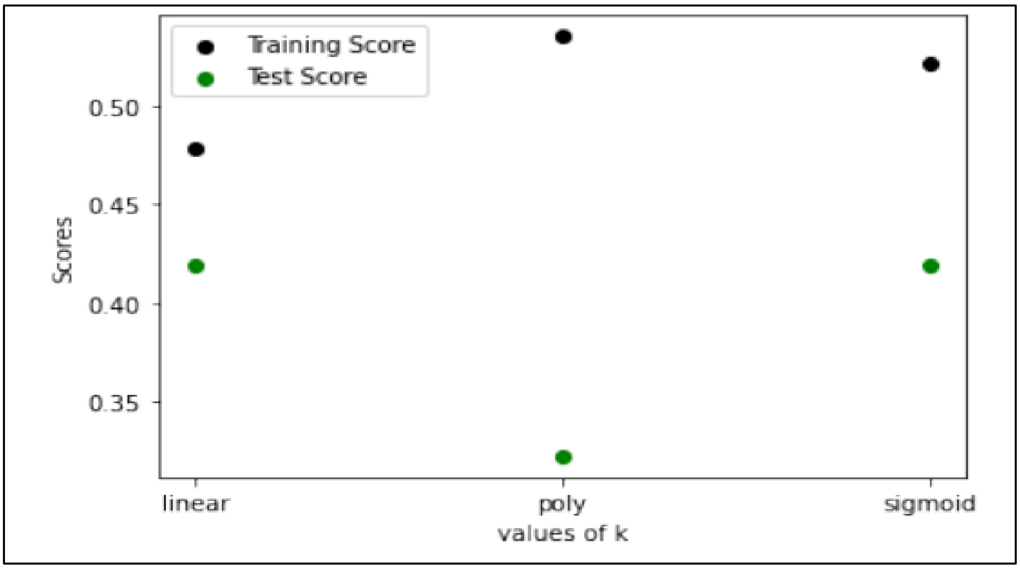


**Figure 11**. “SVM classification scores for male athletes”

In Figure 11, the training score for the three functions is higher than in the test score. We conclude that the generalization of the model is weak for the poly function and average for the other functions. The generalization of the SVM model for female athletes is more beneficial than for male athletes.

- *MLP classifier:*

The MLP classifier implements the multi-layer perceptron algorithm which uses the back propagation method. The MLP layer consists of several parameters like the number of hidden neurons, the layers (input, output and hidden) and the iterations. Regularization: The MLPRegressor and MLPClassifier classes use Alpha parameters for regularization which helps to avoid excess or “overfitting” by penalizing weights with large magnitudes. Women's sport prediction: There are 8 sport classes in the database: 'Swim', 'Row', 'B_Ball', 'T_400m', 'Field', 'T_Sprnt', 'W_Polo' and 'Tennis'. We build three classifiers MLP1, MLP2 and MLP3 named “adam”, “constant learning-rate” and “inv-scaling with Nesterov's momentum”. For the prediction of sport type in female athletes, the following curve explains the result of training scores and testing scores.

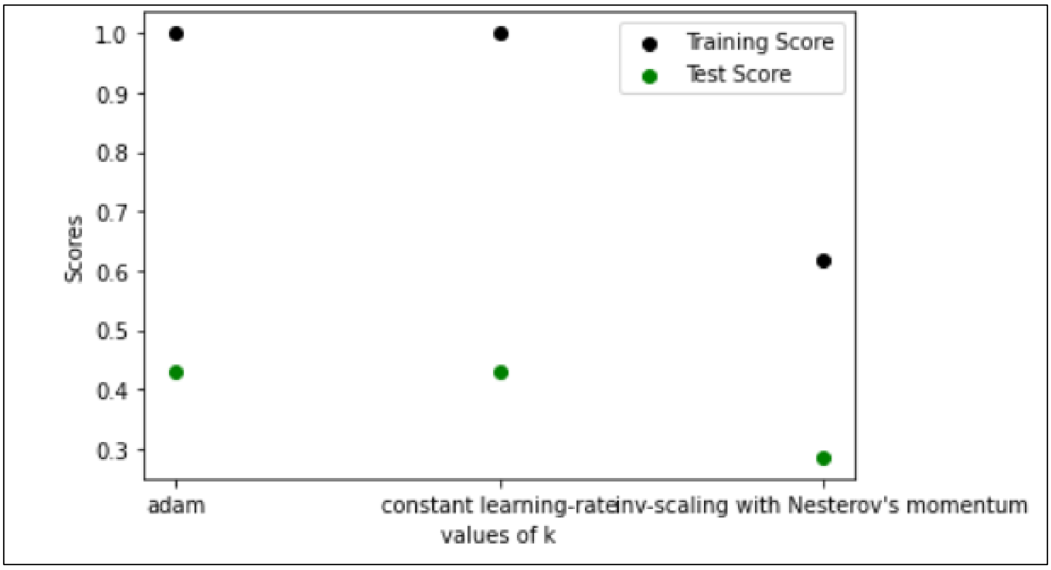


**Figure 12**. “MLP Rating for Female Athletes”

In Figure 12, we distinguish that the MLP3 classifier has a better score than for the MLP1 and MLP 2 classifiers. For the prediction of sport type in male athletes, the following curve explains the result of training scores and testing scores.

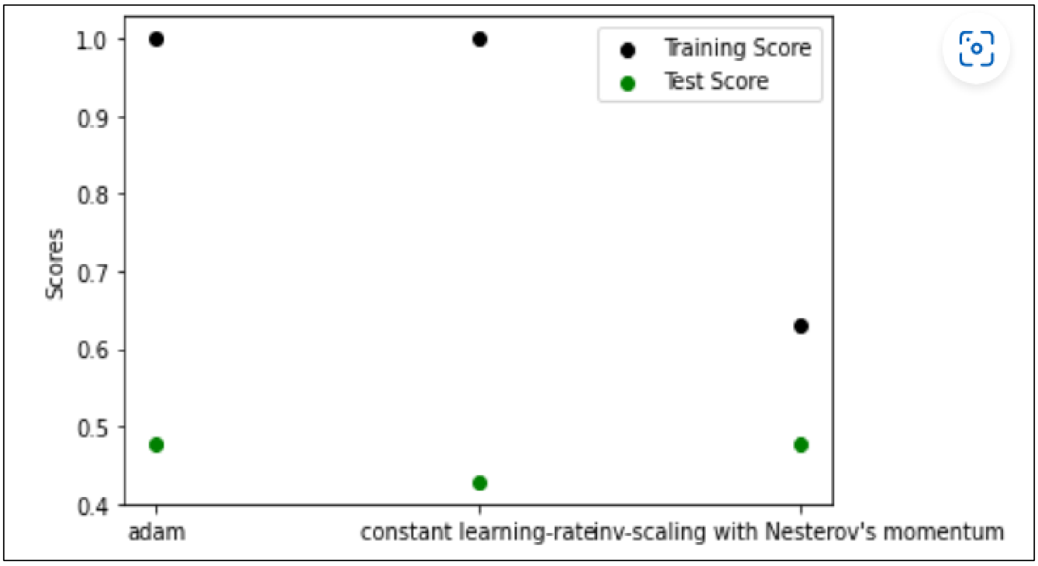


**Figure 13**. "MLP Rating for Male Athletes"

According to Figure 13, we distinguish in this curve that the MLP3 classifier has a better score than for MLP1 and MLP2. The MLP rating for female athletes is better than for male athletes.

*Comparison of classifiers:*

In Table 6, it is concluded from the evaluations of the algorithms that the best evaluation result is that of the SVM algorithm because the test and training scores are closer and as high as in the other classifiers and more precisely for female athletes.

**Table 6.** "Evaluation of classifiers"

| Classifiers | Feminine athletes | Male athletes |
|---|---|---|
| KNN (Generalization) | Between 50% and 60% | Between 25% and 55% |
| SVM (Generalization) | Between 45% and 55% | Between 45% and 55% |
| MLP (Generalization) | Between 30% and 60% | Between 50% and 60% |

**Conclusion:**

The AI has treated many approaches and solutions to best life training for sport players even in existence of many difficulties like accidents and pandemics. ACP method is defined as a reducing technic for data employed in sports to identify key indicators of performance. Many articles and surveys have employed ACP methods in sports actions (Tennis strokes), psychological violence factors and biomechanical movements but none was employed in medicine and science in sports and exercises like our approach. The results of approach interpretation were different from an algorithm to another; results in female athletes really dissimilar with male athletes. Finally, best results in female athletes were in use of KNN algorithm but in male athletes were in use of MLP algorithm.

In future work, learning models will be constructed to do predictive analytics and then this approach will be involved in an AI device that will facilitate the choice of sports for athletes by introducing blood variables, height, weight, and gender. Furthermore, we will supplement the characteristic named “performance” into the dataset to introduce the percentage of the performance of each athlete in practicing sport. Thus, each sportive can be safe from accidents when its performance’s percentage is in a high level.

**References:**

[1] Sports Injury Statistics, https://www.hopkinsmedicine.org/health/conditions-and-diseases/sports-injuries/sports-injury-statistics

[2] Beal, R., Norman, T. J., and Ramchurn, S. D., “Artificial intelligence for team sports: a survey”, Knowl. Eng. Rev. 34., 2019.

[3] Ashley Ying-Ying Wong a, Samuel Ka-Kin Ling, Lobo Hung-Tak Louie, George Ying-Kan Law, Raymond Chi-Hung So et al., “Impact of the COVID-19 pandemic on sports and exercises”, Asia-Pacific Journal of Sports Medicine, Arthroscopy, Rehabilitation and Technology 22 39-44, 2020.

[4] Sergei Bezobrazov, Andrei Sheleh, Sergei Kislyuk, Vladimir Golovko, Anatoliy Sachenko et al., “Artificial Intelligence for Sport Activity Recognition”, the 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Metz, France, 18-21 September 2019.

[5] Alen Rajsp and Iztok Fister, Jr, “A systematic literature review of intelligent data analysis methods for smart sport training”, Appl.Sci, 2020

[6] Sahil Schaudhary and Archana Singh, “Artificial intelligence in sports”, International Conference on contemporary Computing and Informatics (IC3I), 2019.

[7] Heena Rathore, Amr Mohamed, Mohsen Guizani, Shailendra Rathore, “Neuro-fuzzy analytics in athlete development (NueroFATH): a machine learning approach”, Neural Computing and Applications, 2021.

[8] Ana Cristina Nunes Rodrigues, Alexandre Santos Pereira, Rui Manuel Sousa Mendes, André Gonçalves Araújo, Micael Santos Couceiro and António José Figueiredo, “Using artificial intelligence for pattern recognition in a sport context”, Sensors, 2020.

[9] Zhuo Yang, “Research on basketball player’s training strategy based on artificial intelligence technology”, Journal of physics conference series, BTCS 2020.

[10] Nguyen Hoang Nguyena , Duy Thien An Nguyenb , Bingkun Maa and Jiang Hu, “The application of machine learning and deep learning in sport: predicting NBA players’ performance and popularity”, journal of information and telecommunication, 2022.

[11] Arie-Willem de Leeuw, Stephan van der Zwaard, Rick van Baar & Arno Knobbe, “Personalized machine learning approach to injury monitoring in elite volleyball players”, european journal of sport science vol. 22, 2022.

[12] Prem N. Ramkumar, yz MD, MBA, Bryan C. Luu,y BS, Heather S. Haeberley, et al., “Sports Medicine and Artificial Intelligence”, The American Journal of Sports Medicine, 2021.