RBNE
Revista Brasileira de Nutrição Esportiva

MACHINE LEARNING CLASSIFICATION OF CONSUMPTION HABITS OF CREATINE SUPPLEMENTS IN GYM GOERS

Patrícia C. Magalhães[1], Samuel Encarnação[2,3,4,5], André C. Schneider[2], Pedro Forte[2,4,5], José E. Teixeira[2,5,6,7,8], António Miguel Monteiro[2,5], Tiago M. Barbosa[2,5], Ana M. Pereira[5]

ABSTRACT

The aim is to identify usage patterns and the main factors that influence creatine supplementation, providing a basis for future educational interventions and recommendations for safe and effective use. The study was applied to gym goers in Bragança, where a QR code for a survey was released. 158 people participated, 65 non-consumers of creatine supplementation (37.34% men; 22.78% women) and 95 consumers (15.19% men; 24.68% women). Five machine learning algorithms were implemented to classify creatine consumption in gym goers: Logistic Regression, Gradient Boosting Classifier, Ada Boost Classifier, Xgboost Classifier. K-folds cross-validation was implemented to validate the machine learning performance. There was an increased proportion of females with considered themselves not sufficiently informed about the creatine effects/side effects (22.2%) in comparison to males (8.47%), p=0.03. The AdaBoost classifier exposed the best overall performance (86%) in classifying overuse of creatine in gym goers based on their Smoke habits (r = 0.33), grams of creatine used per day (r = 0.50) and lack information about the side effects of creatine intake (r = -0.33). The K-folds method validates the results with very good performance (86%). In conclusion, the five machine learning methods employed well characterized the overuse of creatine in gym goers based on smoke habits, grams of creatine per day, and lack information about the side effects of creatine intake.

Key words: Creatine supplementation. Gyms. Characteristics. Adults.

1 - Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal.
2 - Department of Sports Sciences, Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal.
3 - Department of Physical Education, Sport and Human Movement, Universidad Autónoma de Madrid (UAM), Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain.

RESUMO

Classificação por machine learning dos hábitos de consumo de suplementos de creatina em frequentadores de ginásios

O objetivo é identificar padrões de uso e os principais fatores que influenciam a suplementação de creatina, fornecendo uma base para futuras intervenções educacionais e recomendações para um uso seguro e eficaz. O estudo foi aplicado a frequentadores de academia em Bragança, onde um código QR para uma pesquisa foi disponibilizado. 158 pessoas participaram, 65 não consumidores de suplementação de creatina (37,34% homens; 22,78% mulheres) e 95 consumidores (15,19% homens; 24,68% mulheres). Cinco algoritmos de aprendizado de máquina foram implementados para classificar o consumo de creatina em frequentadores de academia: Regressão Logística, Classificador de Impulso Gradiente, Classificador de Impulso Ada, Classificador Xgboost. A validação cruzada de K-folds foi implementada para validar o desempenho de aprendizado de máquina. Houve uma proporção maior de mulheres que se consideravam insuficientemente informadas sobre os efeitos/efeitos colaterais da creatina (22,2%) em comparação com homens (8,47%), p=0,03. O classificador AdaBoost expôs o melhor desempenho geral (86%) na classificação do uso excessivo de creatina em frequentadores de academia com base em seus hábitos de fumo (r = 0,33), gramas de creatina usadas por dia (r = 0,50) e falta de informação sobre os efeitos colaterais da ingestão de creatina (r = -0,33). O método K-folds validou os resultados com um desempenho muito bom (86%). Em conclusão, os cinco métodos de aprendizado de máquina empregados caracterizaram bem o uso excessivo de creatina em frequentadores de academia com base em hábitos de fumo, gramas de creatina por dia e falta de informação sobre os efeitos colaterais da ingestão de creatina.

Palavras-chave: Suplementação de creatina. Academia. Caracteristicas. Adultos.

INTRODUCTION

Creatine is a compound derived from three amino acids, being non-essential precisely because there is endogenous synthesis, a member of the family of phosphagen guanidines (Kreider et al., 2017).

It plays an essential role in energy metabolism, through a reversible reaction catalyzed by the enzyme creatine kinase (CK), creatine is phosphorylated and converted into phosphocreatine.

This, in turn, is subsequently hydrolyzed, providing a phosphate group to the molecule adenosine diphosphate (ADP), and enough energy to form a new molecule of adenosine triphosphate (ATP), being produced endogenously or acquired through diet (Brosnan, Brosnan, 2016; Wyss, Kaddurah-Daouk, 2000).

The main dietary sources are meat and fish (Kreider et al., 2017). Creatine is predominantly stored in skeletal muscle (95%) and in smaller amounts in the brain and kidneys (5%) (Buford et al., 2007; Kreider; Jung, 2011).

Creatine is stored in skeletal muscle in the form of free creatine and phosphocreatine, which through phosphorylation mechanisms allows for the resynthesis of ATP (Jäger et al., 2017). High-energy phosphates play a vital role in various metabolic processes, especially in the first 10 seconds of high-intensity exercise, with phosphocreatine depletion being one of the main mechanisms causing fatigue (Jäger et al., 2017).

Approximately two-thirds of intramuscular creatine is phosphocreatine (PCr), and the remaining is free creatine. The total of creatine and phosphocreatine (PCr + Cr) in muscle averages 120 mmol/kg of dry muscle mass for a 70kg individual, however, in most individuals, the upper limit for creatine storage is 160 mmol/kg of dry muscle mass (Kreider et al., 2017).

Due to the breakdown of creatine into creatinine and its subsequent excretion, the body needs to compensate for this loss through dietary intake and endogenous synthesis of Cr, at around 1-3g per day, varying according to the individual's levels of muscle mass (Kreider et al., 2017).

Humans can resynthesize creatine efficiently for the normal functioning of the body. The authors mention the value of 2g as the amount needed for a 70kg individual to replenish their creatine reserves (Brosnan, Brosnan, 2016).

Creatine is transported in the bloodstream and absorbed from the intestinal wall, and then directed to target organs, such as skeletal muscle, brain, heart, and testicles (Balestrino, Adriano, 2019). Creatine synthesis primarily occurs in the liver and kidneys, and the brain seems to be able to synthesize its own creatine (Braissant et al., 2011; Brosnan, Brosnan, 2016). The synthesis process involves essentially two enzymes, arginine: glycine amidinotransferase (AGAT) and guanidinoacetate methyltransferase (GAMT) enzyme. AGAT is present in the kidneys, brain, and pancreas, and condenses the amino acids arginine and glycine, forming guanidinoacetate methyltransferase (GAA).

Using the enzymatic cofactor S-adenosylmethionine (SAM), the guanidinoacetate N-methyltransferase enzyme transfers a methyl group to GAA, thus forming creatine. Any errors in AGAT, GAMT, and CrTr (creatinine transporter) compromise the synthesis of creatine (Brosnan, Brosnan, 2016).

Creatine monohydrate is the most studied type of creatine and is also considered the most effective compared to other chemical forms. The uptake of creatine involves the absorption of creatine into the blood and transportation to target tissues, with the peak concentration of creatine in the blood appearing to occur 60 minutes after consumption (Jäger et al., 2017; Kreider et al., 2017).

The supplementation of this compound appears to be able to increase intramuscular creatine levels in individuals of various age groups, from children to the elderly (Kreider et al., 2017).

It is one of the most popular ergogenic supplements for athletes, with a solid body of knowledge on the intramuscular increase in creatine and its effect on improving performance and athletic performance, with high-intensity performance increasing by 10 to 20%.

The ergogenic value of creatine, applied to various sports modalities, presents itself through increased performance in single or repeated sprints, increased performance in maximum strength training repetitions, increased muscle mass and strength (as an adaptive response to strength training), increased glycogen synthesis, improved aerobic capacity, improved muscle recovery capacity, and better tolerance to training (Kreider, Jung, 2011).

Creatine is stored in skeletal muscle, in the form of free creatine and phosphocreatine, which through phosphorylation mechanisms allows for the resynthesis of ATP (Jäger et al., 2017).

The daily intake of creatine is about 1 to 2g/day, saturating creatine stores by approximately 60 to 80% (Kreider et al., 2017).

Supplementation causes an increase in muscle phosphocreatine levels of about 20 to 40%, depending on basal levels (Jäger et al., 2017).

The best way to saturate creatine reserves is to ingest 5g (approximately 0.3g/kg body weight), four times a day, for 5-7 days, usually called the "loading phase" (Harris, Söderlund, Hultman, 1992; Hultman et al., 1996).

Once creatine reserves are saturated, maintenance can be achieved with 3-5g/day of creatine, although larger athletes may need 5-10g/day for maintenance (Kreider et al., 2017).

Alternatively, a 3g/day protocol for 28 days seems to result in similar outcomes in terms of cellular phosphocreatine levels compared to the initial creatine saturation process (Hultman et al., 1996).

The safety of creatine supplements is well established in both adult athletes and non-athletes through short and long-term studies (up to 5 years) that show no alterations in clinical health indicators (Kreider et al., 2010).

Furthermore, creatine is also not associated with human carcinogenesis or the formation of heterocyclic amines. The only side effect caused by creatine can be understood as an increase in body weight of approximately 1.6 to 2.4kg, which can only be harmful in speed sports where it is necessary to control body weight for better athletic performance (Calfee, Fadale, 2006).

There are mixed findings in the literature, such as studies that associate the exogenous ingestion of creatine with symptoms such as mild gastro-intestinal discomfort, nausea, and cramps (Calfee, Fadale, 2006).

However, the position of the International Society of Sports Nutrition is that all studies conducted to date, involving population groups ranging from children to the elderly, with administration of creatine doses between 0.3 to 0.8g/kg/day for up to 5 years, have not shown any risk of compromising health due to creatine supplementation.

Creatine supplementation is a common practice among athletes and gym-goers to improve physical performance. With the increasing number of gym-goers, many myths and misinformation about its use also arise, which can lead to serious health problems due to the improper use of creatine (Gualano et al., 2010).

Although creatine is generally considered safe when used according to recommendations, it is paramount referring to a healthcare professional before starting any supplementation, especially if there is a history of health issues or the use of other medications (Kreider et al., 2017).

The aim is to identify usage patterns and, through machine learning algorithms, classifying the main factors that influence creatine supplementation and its overuse level by the gym goers, providing a basis for future educational interventions and recommendations for safe and effective use.

We believe that the level of knowledge about creatine supplementation significantly influences the amount of creatine used by gym-goers.

Additionally, demographic factors such as age, sex, training experience level, and lifestyle habits may be associated with variations in creatine use.

Finally, we consider that the frequency and intensity of training are positively correlated with the amount of creatine used by weightlifters.

MATERIALS AND METHODS

SAMPLE

Simple sampling technique was employed, and 158 gym-goers were recruited. All participants met the inclusion criteria, namely: being over 18 years old, being a gym-goer in Bragança, and agreeing to the conditions established initially for participation in the study.

Table II shows a comparison of the percentage values of the sample for both sexes and for non-users and users of creatine. No differences were noted male and female groups ($p>0.05$), revealing that the data were homogeneous for conducting the research.

Data Collection Instruments

The study was applied to gym-goers in the city of Bragança during of September 2023. A QR code of an electronic survey was disseminated, conducted through the online Google Forms platform, in gyms. Regarding the questionnaire, it includes three sections. The first and second sections refer to sociodemographic details (nationality, sex, age, education level) and self-reported anthropometric data (weight and height) used to calculate BMI, respectively.

The third part is an adaptation of a questionnaire by Martins (2022), concerning the consumption of creatine supplementation, including questions regarding the number and frequency of workouts, type of supplement consumption per workout, number of doses, goals, or reasons for consumption.

Procedures

Several individuals were approached at gyms in the city of Bragança, and those who met the inclusion criteria through a QR code were given access to the questionnaire. This includes an informed consent form, and only by agreeing to all defined conditions were they allowed to answer the questions.

After collecting data from all participants, a table was formulated, and the data was organized by sex and by the condition of consuming or not consuming creatine supplementation.

Ethical procedures

The study protocol was submitted and approved by the Ethics Committee of the Polytechnic Institute of Bragança (Process Nº. 50385).

All procedures followed the recommendations of the Helsinki Declaration and the Oviedo Convention (Helsinki Declaration of the World Medical Association Ethical Principles for Medical Research Involving Human Subjects, n.d.) and participants agreed to the terms of informed consent for their participation in the study.

Descriptive analysis

In this study, statistical analysis was conducted for sample description and comparative analysis between groups by sex.

Data was described in absolute and percentual values. For this purpose, we applied a significance test for two proportions from unpaired sample, suggested by Gardner et al. (1989); Altman, (2011). on their 6th chapter. This proportional test is based on the null hypothesis that points that two proportions are equal based, and an alternative hypothesis which states that the two proportions differ between its selves [based on interval of confidence of 95% ($p<0.05$) (Altman, 2011).

We respected the assumption of Gardner et al., (1989) which reports that the proportions must be arranged in a range of 0.1-0.9. In this way, this test calculates the Z-score by dividing the difference of proportions between the two sample (prop 1 - prop 2) minus the standard error calculated be the two sample proportions. By the Z score, we verified the p-values in the table of p-values available in the Chapter 6 (Altman, 2011).

We performed all statistical analyses by sing the PythonTM programming language (Naranjo et al., 2022).

Classification Algorithm implementation

To identify the use of creatine among gym goers, we considered the following group as independent/predictor variables: 1st being a foreigner, 2nd sex, 3rd age, 4th education level, 5th body weight, 6th height, 7th BMI, 8th having a disease, 9th consuming alcoholic beverages, 10th being a smoker, 11th number of weekly training sessions, 12th duration of training (hours), 13th amount of creatine ingested per day (grams), 14th time of creatine ingestion, 15th level of information regarding creatine supplementation, , 16th reading the label of the creatine packaging to be used, 17th having knowledge of the benefits, 18th having knowledge of the side effects, 19th source of information that led the participant to start using creatine, 20th having a motivational factor for the use of creatine, 21st having two motivational factors for the use of creatine, and 22nd having three or more motivational factors for the use of creatine. Following the suggestions of Kumar, Chong, (2018). The best selection of predictor variables was made through a correlation matrix between all binary variables (0 or 1). To verify the statistical significance of the correlation matrix, cutoff points for effect size (TE) were adopted, with Cohen's $r \leq 0.10$ = small, ≤ 0.30 = moderate, and ≥ 0.50 = large (Cohen, 2013).

For variables observed as multiclass (more than two categories), a one-hot-encoding process was carried out at a single point, thus assigning characteristics attributed as binary (0 or 1) for each category (Henry et al., 2015).

For both processes of selecting the best variables, two dependent variables of interest were considered, the 1st being the amount of creatine ingested per day (grams) and the 2nd being excessive daily creatine use. After selecting the best variables, it was noted that only the following set of independent variables [1st sex, 2nd BMI, 3rd being a smoker, 4th amount of creatine ingested per day (grams), 5th excessive use, 6th level of information regarding creatine supplementation, 7th knowledge of side effects] presented the best correlations when related to the variable "excessive daily creatine use," Table I. Nevertheless, only the variables “Being a smoker” (moderate ES, r = 0.33), “Grams of creatine consumed per day” (Large ES, r = 0.50), and “Knowledge about the side effects of the supplement” (moderate ES, r = -0.30), and these variables were included in the final panel of predictors of the gym goers’ creatine overuse.

Table 1 - Final relationship of the best variables and gym goers’ creatine overuse.

Independent Variables	Target Variable
Sex	0.14
BMI	0.24
Being a smoker	0.33
Grams of creatine consumed per day	0.50
Level of information about the suppl(Jäger et al., 2017) ement	- 0.21
Knowledge about the side effects of the supplement	-0.30

Note: Larger effects (selected predictors are highlighted in bold).

In this way, due to the binary characteristic of the target variable (gym goers’ creatine overuse), we choose to perform machine learning algorithms for classification tasks (Chicco, Jurman, 2020).

Thus, to classify the Gym goers’ creatine consumption, using PythonTM, programming language, and from the main library “sklearn”, we implemented four machine learning algorithms: 1º Logistic regression classifier, activating the function “from sklearn. linear_model import Logistic Regression”, 2º Gradient Boosting classifier, activating the function “from sklearn. ensemble import Gradient Boosting Classifier”, 3º Ada Boost classifier, activating the library “from sklearn. ensemble import Ada Boost Classifier”, and 4º Extreme Gradient Boosting classifier, using the function “from xgboost import XGB Classifier” (Raschka, Patterson, Nolet, 2020).

Thus, the best variables (Smoke, grams of creatine used per day, and information about the side effects of creatine intake) were allocated in array X (predictive array) and the variable "Gym goers’ overuse" was allocated within array Y (dependent variable) (Haun et al., 2019). In the next step, the data set was distributed 30% for training and 70% for test, using a random seed = 0 (Liu, Cocea, 2017). To evaluate the model’s overall learning performance, mean accuracy, explained as accuracy = true positives + true negatives / true positives + true negatives + false positives + false negatives (Fawcett, 2006).

For the evaluation of the amount of correctly made positive predictions in the model, the precision equation was applied, understood as Precision = true positives / true positives + false positives. For the evaluation of positive predictions about the number of false negatives, recall was applied, understood as recall = true positives / true positives + false negatives. Finally, for evaluating the model's precision among class imbalances, the F-1 score was applied, understood as F-1 score = 2 x precision x recall/precision + recall (Chicco, Jurman, 2020). Performance metrics and a confusion matrix (number of false positives and false negatives) of the model were displayed in the table 6.

Cross-validation of the classifying algorithms

A K-fold cross-validation was applied to validate the models’ performance in classifying the gym goers’ creatine consumption. This method takes the original dataset and builds K subsets from it, training and testing the algorithms again, but in each subset, thus, seeing the algorithms' instability in classifying outcomes throughout different parts of the original dataset, thus, proving the homogeneity of the results (Cunha, 2019).

For this validation, we selected two set ups of folds, being five or 10 folds from the original X array used in the algorithms training and testing procedures and we choose those folds setting up that presented the best validation performance (Cunha, 2019).

RESULTS

Demographics data

The Table 2 describes the proportion analysis for the participants' demographic information. There was only a significant difference for the male group of non-users of creatine, which presented lower incidence of chronic diseases compared with the female group, p = 0.01.

Table 2 - Sample demographic.

Variables	Non-creatine consumers				Creatine consumers			
	Mal. (n=24, 15.19%)	Fem. (n=39, 24.68%)	Z	p	Mal. (n=59, 37.34%)	Fem. (n=36, 22.78%)	Z	p
Years								
18 to 25	13 (54.17)	18 (46.15)	0.43	0.69	32 (54.24)	19 (52.78)	0.10	0.92
26 to 30	4 (16.67)	12 (30.77)	-0.54	1.00	14 (23.73)	8 (22.22)	0.08	0.92
31 to 35	3 (12.50)	3 (7.69)	-1.59	1.00	7 (11.86)	2 (5.56)	-1.32	1.00
36 to 45	2 (8.33)	3 (7.69)	0.08	0.92	5 (8.47)	5 (13.89)	-1.32	1.00
over 45	2 (8.33)	3 (7.69)	0.08	0.92	1 (1.69)	2 (5.56)	-0.63	1.00
Education								
No education	0 (0.0)	0 (0.0)	NA	NA	0 (0.0)	1 (2.78)	NA	NA
Grade 9	1 (4.17)	0 (0.0)	NA	NA	1 (1.69)	0 (0.0)	NA	NA
High school	5 (20.83)	2 (5.13)	-0.82	1.00	16 (27.12)	7 (19.44)	0.39	0.76
Bachelor's degree	9 (37.50)	19 (48.72)	-0.58	1.00	25 (42.37)	11 (30.56)	0.66	0.55
Master's degree	8 (33.33)	13 (33.73)	-0.01	1.00	14 (23.73)	12 (33.33)	-0.54	1.00
PhD	1 (4.17)	3 (7.69)	-0.65	1.00	1 (1.69)	2 (5.56)	-0.63	1.00
Postgraduate	0 (0.0)	2 (5.13)	NA	NA	2 (3.39)	2 (5.56)	-0.43	1.00
Pos-secondary vocational programme	0 (0.0)	0 (0.0)	NA	NA	0 (0.0)	1 (2.78)	NA	NA
Consumes alcoholic drinks								
Every day	1 (4.17)	0 (0.0)	NA	NA	1 (1.69)	0 (0.0)	NA	NA
At least once a week	13 (54.17)	13 (33.33)	1.07	0.31	18 (30.51)	13 (36.11)	-0.32	1.00
At least once a month	5 (20.83)	11 (28.21)	-0.31	1.00	18 (30.51)	11 (30.56)	-0.002	1.00
No	5 (20.83)	15 (38.46)	-0.72	1.00	22 (37.29)	12 (33.33)	0.23	0.84
Smoker								
Yes	4 (16.67)	8 (20.51)	-0.15	1.00	12 (20.34)	10 (27.78)	-0.40	1.00
No	20 (83.33)	31 (79.49)	0.34	0.76	47 (79.66)	26 (72.22)	0.72	0.92
BMI								
Underweight	1 (4.17)	1 (2.56)	0.24	0.84	0 (0.0)	1 (2.78)	NA	NA
Normal weight	11 (45.83)	30 (76.92)	-1.90	1.00	34 (57.63)	28 (77.78)	-1.67	1.00
Overweight	12 (50)	8 (20.51)	1.61	0.92	23 (38.98)	6 (16.67)	1.02	0.23
Obesity grade 1	0 (0.0)	0 (0.0)	NA	NA	2 (3.39)	1 (2.78)	0.10	0.92

Chronic Diseases								
Yes	2 (8.33)	7 (17.95)	1.66	0.10	1 (1.69)	6 (16.67)	0.05	0.92
No	22 (91.67)	32 (82.05)	1.00	0.31	58 (98.31)	30 (83.33)	2.64	0.01

Note: NA: not applicable. The significant values are marked in bold.

Training program and supplementation habits

The table 3 represents the sample training program and supplementation habits. After the proportion test calculation, there were no significant differences between groups. Despite the lack of statistical differences between groups, the male group (25.42%) and the female group (13.89%) presented consumption levels of creatine above the recommended amount in the products labels. The declared portion on the nutritional labeling is the minimum daily amount recommended by the manufacturer (3000 mg) for adults (≥19 years) (Araújo, Amorim, Silva, 2O23).

Table 3 - Sample training program and supplementation habits.

Variables	Mal. (n=59, 37.34%)	Fem. (n=36, 22.78%%)	Z	p
Sess. Week.				
1 to 2	1 (1.69)	0 (0)	NS	NS
3 to 4	11 (18.64)	8 (22.2)	-0.21	1.00
4 to 5	31 (52.54)	20 (55.56)	-0.27	1.00
More than 5	16 (27.12)	8 (22.2)	0.26	0.84
Training hours				
1	22 (37.29)	15 (41.67)	-0.24	1.00
2	35 (59.32)	21 (58.33)	0.07	0.92
3	1 (1.69)	0 (0)	NS	NS
4	1 (1.69)	0 (0)	NS	NS
Amount of creatine ingested per day				
< 3	3 (5.08)	13 (36.11)	0.44	0.69
3 to 5	34 (57.62)	21 (58.33)	1.28	0.23
5 to 10	20 (33.89)	2 (5.56)	-0.45	1.00
10 to 20	2 (3.39)	0 (0)	NS	NS
Consumes more than the recommended amount on the label				
Yes	15 (25.42)	5 (13.89)	0.53	0.62
No	44 (74.58)	31 (86.11)	-1.21	1.00
Consumption is done				
Before training	26 (44.07)	21 (58.33)	-0.97	1.00
During training	1 (1.69)	1 (2.78)	-0.18	1.00
After training	27 (45.76)	10 (27.78)	0.98	0.37
Before and after training	5 (8.47)	4 (11.11)	-2.19	1.00

Note: NS: not significant. Mal.: Male, Fem.: Female.

Information and knowledge

The table 4 shows the characteristics related to the level of information and knowledge about the creatine supplementation in gym goers. There was an increased proportion of females with considered themselves not sufficiently informed about the creatine effects/side effects (22.2%) in comparison to males (8.47%), p = 0.03.

Table 4 - Characteristics of information and knowledge of the gym goers that supplement creatine.

Variables	Masc. (n=59, 37.34%)	Fem. (n=36, 22.78%)	Z	p
Informed about the use of supplements				
Yes	54 (91.53)	28 (77.78)	1.74	0.08
No	5 (8.47)	8 (22.2)	2.18	0.03
Reads the label				
Yes	51 (86.44)	32 (88.89)	-0.39	1.00
No	8 (13.56)	4 (11.11)	0.09	0.92
Do you know the benefits				
Yes	59 (100)	36 (100)	0.00	1.00
No	0 (0)	0 (0)	0.00	1.00
Do you know the side effects				
Yes	47 (79.66)	25 (69.44)	0.93	0.37
No	12 (20.34)	11 (30.56)	-2.79	1.00

Note: The significant values are marked in bold

Motivation for consumption

The table 5 represents the characteristic of the factors that are related to the participants' creatine use. The test of proportions did not identify any significant difference between groups.

Table 5 - Characteristics of the factors that are related to the participants' creatine use.

Variables	Masc. (n=59, 37.34%)	Fem. (n=36, 22.78%)	Z	p
Source of information/counseling				
Own initiative	40 (67.80)	24 (66.67)	0.08	0.92
Personal trainer	13 (22.03)	6 (16.67)	0.25	0.84
Nutritionist	11 (18.64)	6 (16.67)	0.10	0.92
Doctor	4 (6.67)	1 (2.78)	0.71	0.48
Friends	12 (20.34)	4 (11.11)	0.82	0.42
Family members	5 (8.47)	0 (0)	NA	NA
Supplement vendors	3 (5.08)	3 (8.33)	-0.77	1.00
Social media	19 (32.20)	10 (27.78)	0.27	0.84
Online research	2 (3.39)	1 (2.78)	0.05	0.92
Objectives/Motives				
Increase energy	33 (55.93)	25 (69.44)	-1.08	1.00
Increase strength	43 (72.88)	26 (72.22)	0.09	0.92
Gain muscle mass	32 (54.24)	22 (61.11)	-0.51	1.00
Lose fat mass	7 (11.86)	6 (16.67)	-0.20	1.00
Reduce recovery time between workouts	24 (40.68)	16 (44.44)	-0.24	1.00
Decrease pain after workout	12 (20.34)	9 (25)	-0.27	1.00
Decrease fatigue after workout	15 (25.42)	9 (25)	0.02	0.92
Reduce stress	5 (8.47)	5 (13.89)	-2.23	1.00
Improve sleep	6 (10.17)	6 (16.67)	-0.33	1.00
Increase muscle endurance	1 (1.69)	0 (0)	NS	NS

Note: NS: not significant.

Results of the machine learning classifying models

The table 6 displays the results of the machine classifying models applied. The AdaBoost classifier exposed the best performance (average performance = 0.86) in classifying the gym goers overuse of creatine.

Table 6 - Outputs of machine learning algorithms implemented.

Algorithm	Acc.	Prec.	Rec.	F-1 Score	AUC	x̄	TN	FP	FN	TP
Logistic Regression	0.89	0.60	0.75	0.66	0.95	0.77	23	2	1	3
Gradient Boosting Classifier	0.79	0.87	0.79	0.81	0.78	0.80	20	5	1	3
Ada Boost Classifier	0.86	0.89	0.86	0.87	0.81	0.86	22	3	1	3
Xgboost Classifier	0.83	0.88	0.83	0.85	0.79	0.84	21	4	1	3
Overall Performance (x̄)	0.84	0.81	0.79	0.79	0.83	0.77	NA	NA	NA	NA

Note: Acc.: accuracy, Prec.: precision, Rec.: recall, AUC: area under the curve, x̄: arithmetical average, TN: true negative, FP: false positive, FN: false negative, TP: true positive. NA: not applicable.

Results of the cross-validation model

The table 7 represents the K-folds cross-validation results. The results indicated that the algorithms presented a very similar performance when divided into five folds (primary algorithm results = 81.2% vs k-fold cross-validation = 81%).

Table 7 - Outputs of the cross-validation of the classifying models' performance.

Algorithm	Accuracy (%)	Accuracy (Sub.1)	Accuracy (Sub.2)	Accuracy (Sub.3)	Accuracy (Sub.4)	SD
Logistic Regression	0.79	0.74	0.74	0.74	0.94	0.78
Gradient Boosting Classifier	0.79	0.74	0.79	0.79	0.84	0.79
Ada Boost Classifier	0.79	0.79	0.68	0.84	0.84	0.79
Xgboost Classifier	0.82	0.79	0.79	0.79	0.84	0.89
Overall Performance (x̄)	0.80	0.77	0.75	0.79	0.87	0.81

Note: Sub.: subsets from the entire X array. SD: standard deviation.

DISCUSSION

The present study aimed to identify usage patterns and, through machine learning algorithms, classifying the main factors that influence creatine supplementation and its overuse level by the gym goers.

The data from this study shows that approximately 60.1% of participants consuming creatine. The main findings of this study were that the variables being a smoker, the amount of creatine per day, and the level of knowledge about the side effects of the creatine use were the most important features in predicting the gym goers' creatine overuse levels. It was proved by the good performance from the classifying algorithms (81.2 %), that was well-validate by the K-folds cross-validation test with a vary closed performance (81%).

Another important component was grams of creatine consumed per day, that presented a positive correlation with large effect size (r=0.50) with the gym goers' creatin overuse. In addition, both groups presented higher percentages of creatine use in the age category of 18-25 years old for both sexes, and there were 25.42% of males and 13.89% of females intaking amounts above the daily recommended.

The combination of high consumption rates among young adults and the significant correlation with excessive use highlights the need for awareness and education strategies to ensure that creatine supplementation is done safely and responsibly.

Although creatine supplementation does not pose significant risks to the health of healthy individuals, there may be impairment of kidney function when use is indiscriminate (more than 5 grams per day).

Therefore, it is essential that users are aware of safe limits and follow appropriate recommendations to avoid health complications (Ataídes, Aguiar Neto Filho, Santos, 2022).

The literature notes that creatine should not be used for prolonged periods, suggesting a three-month supplementation interval (Peralta, Amancio, 2002).

Continuous consumption for up to three-months was also reported in other studies (Garrido et al., 2008; Reis et al., 2017). A study conducted in Campinas-SP, Brazil, showed a percentage of 36.3% of weightlifters aged between 23 and 28 years old. In Caxias do Sul-RS, Brazil, an average age of 22 years old was found among gym-goers (Lollo, 2004; Theodoro, Ricalde, Amaro, 2009). In the present study, over 50% of weightlifters were between 18 and 25 years old, in line with previous studies.

However, we can observe a prevalence of males and found a positive association with a small effect size indicating a possible higher excessive use of creatine in the male group, as in other studies that showed a strong presence of males as supplement users (Frade et al., 2016; Garrido et al., 2008; Lollo, 2004; Reis et al., 2017; Theodoro, Ricalde, Amaro, 2009). There are differences in the storage and use of creatine between healthy men and women (Brosnan, Brosnan, 2007).

Women produce 70 to 90% less creatine than men (2.6-6.2 mmol/day for women; 3.7-7.7 mmol/day for men). The lower consumption in women, as also observed in our study, that compared to men, explains the decreased levels of creatine excretion (0.75 for women and 1 for men) (Ellery et al., 2016). Studies like that of Kreider et al., (2017) found a higher adherence to creatine supplementation among men, 28 to 29%, while in women a value of 0.23 to 0.3% was observed.

These data suggest a sex disparity in the use of supplements, possibly influenced by differences in training goals and perceptions about supplementation (Jagim et al., 2018). Men appear to be more likely than women to report the use of creatine, often motivated by the desire to increase lean body mass and energy production (Jagim et al., 2018).

Authors draw attention to supplement users, emphasizing that when consuming them, considerations should be taken into account regarding diet, workout intensity, and frequency to not overload the body causing adverse effects associated with an excess of nutrients in the body (Garrido et al., 2008; Reis et al., 2017).

In the variable information level about the supplement, 8.47% of men and 22.22% of women who consume creatine supplements do not feel adequately informed about the supplement. These values are very close to the study by Sousa et al., (2008) where 33.7% of Portuguese participants did not feel sufficiently informed about supplements. With the data from our study, we found a positive association with a small effect size indicating a possible higher overuse of creatine by the less informed group. Self-administration of supplements is a reality that has increased with the rise of information on social media. These results are supported by the studies of Costa et al., 2013, and Goston, Mendes 2011). Which suggest that the majority of supplement consumers do not seek advice from healthcare professionals. This fact supports the results of this study, in which those who are less informed have improper use.

In terms of knowledge about the side effects of the supplement, 20.34% of men and 30.56% of women are unaware of the possible adverse effects of creatine supplementation. It was reinforced when we found an inverse correlation between the knowledge about the side effects of the creatine supplementation and its overuse by the gym goers, which was an important feature into the classificatory analysis with machine learning. Complementarily, the proportion analysis showed that the female group (22.2%) was less unformatted about the use of supplements than the male group (8.47%).

The disparity in chronic diseases between men (98.31%) and women (83.33%) without creatine supplementation is the result of complex biological and social factors. Biologically, women have a higher prevalence of autoimmune diseases, depression, and cardiovascular diseases, affected by hormonal fluctuations and genetic predispositions.

Additionally, social factors, such as differences in access to healthcare and inequalities in treatment, may contribute to underdiagnosis and inadequate treatment of these conditions in women (Temkin et al., 2023).

In contrast, men have a lower incidence of these diseases (WHO, 2023), possibly related to differences in immune response and other less understood biological factors.

Although creatine supplementation may offer additional health benefits, such as improved muscle function and potential neuroprotective effects, its use is more common among men, which may influence the observed health outcomes, highlighting the importance of healthcare strategies that recognize and address the specific needs of both sexes.

Although the direct correlation between smoking and excessive creatine use has not been widely investigated in the scientific literature, both habits have potential implications for health. Smoking is known to be harmful to health, significantly affecting the cardiovascular and respiratory systems, while excessive creatine use can carry risks such as kidney damage (Ataídes, Aguiar Neto Filho, Santos, 2022; Varghese, Muntode Gharde, 2023; WHO, 2023). For individuals who engage in both behaviors, it is crucial to carefully consider the overall health impacts and seek specialized medical guidance.

Regarding the reasons why respondents use creatine supplementation, 72.88% of men and 72.22% of women stated that they use it to increase strength. According to Garrido et al., (2008), the main reason for using creatine was to increasing endurance (45%), followed by aesthetic reasons (36%), including hypertrophy and muscle definition. Creatine is used by athletes and gym-goers mainly to increase protein synthesis and muscle mass (Franco et al., 2007), as well as to improve performance in short, high-intensity exercises (Gualano, Artioli, Lancha Junior, 2008).

The American Dietetic Association recommends adopting a balanced diet and seeking professional guidance before incorporating supplements into the diet to prevent outcomes like those in this study, where those who were not aware of the side effects of creatine tend to use it excessively (Rodriguez et al., 2009).

A limitation of this study is that the sample size could have been larger, which may have limited the analysis (only 96 creatine consumers), thus limited the interpretation of reality and possibly brought limitations to the algorithm learning. Other limitations include not having assessed the fitness level of each individual, which could have helped better understand the model, and self-reported weight and height without confirmation.

Additionally, the recruited population was generalized, which may introduce bias to the model, such as individuals with different training times/experiences, considered a strong confounding factor.

Finally, another limitation was the cross-sectional design that prevented us from establishing any causality between determinants and creatine use among participants.

As strengths, we highlight the algorithms' performance in classifying the overuse of creatine among gym goers in an unbalanced dataset where the majority of the creatine users were intaking amounts of creatine into the recommended amounts (74.58% for males and 86.11% for females), showing good replication properties in additional studies.

CONCLUSION

It was found that the variables being a smoker, grams of creatine ingested per day, and knowledge about the side effects of the supplement was the best variables to classify the creatine overuse among the participants (25.42% of males and 13.89% of females).

REFERENCES

1-Altman, D.G Statistics with confidence: confidence intervals and statistical guidelines. 2ª edição. London. BMJ Books. 2011.

2-Araújo, P.L.M.; Amorim, A.G.; Silva, T.A. Análise de rótulos de suplementos alimentares de creatina. Revista Brasileira de Nutrição Esportiva. São paulo. Vol. 17. Num. 106. p. 600-609. 2023.

3-Ataídes, K.C.; Aguiar Neto Filho, M.; Santos, J.D.S.G. Benefícios e malefícios da suplementação com creatina. Scientific Electronic Archives. Vol. 15. Num. 10. 2022.

4-Balestrino, M.; Adriano, E. Beyond sports: Efficacy and safety of creatine supplementation in pathological or paraphysiological conditions of brain and muscle. Medicinal Research Reviews. Vol. 39. Num. 6. 2019. p. 2427-2459.

5-Braissant, O.; et al. Creatine deficiency syndromes and the importance of creatine synthesis in the brain. Amino Acids. Vol. 40. Num. 5. 2011. p. 1315-1324.

6-Brosnan, J.T.; Brosnan, M.E. Creatine: Endogenous Metabolite, Dietary, and Therapeutic Supplement. Annual Review of Nutrition. Vol. 27. 2007. p. 241-261.

7-Brosnan, M.E.; Brosnan, J.T. The role of dietary creatine. Amino Acids. Vol. 48. Num. 8. 2016. p. 1785-1791.

8-Buford, T.W.; et al. International Society of Sports Nutrition position stand: creatine supplementation and exercise. Journal of the International Society of Sports Nutrition. Vol. 4. 2007. p. 6.

9-Calfee, R.; Fadale, P. Popular ergogenic drugs and supplements in young athletes. Pediatrics. Vol. 117. Num. 3. 2006. p. e577-589.

10-Chicco, D.; Jurman, G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making. Vol. 20. Num. 1. 2020. p. 16.

11-Cohen, J. Statistical Power Analysis for the Behavioral Sciences. Routledge. 2013.

12-Costa, D.C.; Rocha, N.C.A.; Quintao, D.F. Prevalence of use of dietary supplements between practitioners of physical activity in academies of two cities in the Vale do Aco-MG: associated factors. Revista Brasileira de Nutrição Esportiva. São Paulo. Vol. 7. Num. 41. 2013. p. 287-300.

13-Cunha, J.P.Z. Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos. Mestrado em Estatística. São Paulo: Universidade de São Paulo. 2019.

14-Ellery, S.J.; et al. Dietary creatine supplementation during pregnancy: a study on the effects of creatine supplementation on creatine homeostasis and renal excretory function in spiny mice. Amino Acids. Vol. 48. Num. 8. 2016. p. 1819-1830.

15-Fawcett, T. Introduction to ROC analysis. Pattern Recognition Letters. Vol. 27.2006. p. 861-874.

16-Frade, R.E.; et al. Avaliação do consumo de suplementos nutricionais por frequentadores de uma academia da cidade de São Paulo-SP. Revista Brasileira de Nutrição Esportiva. São Paulo. Vol. 10. Num. 55. 2016. p. 50-58.

17-Franco, F.S.C.; et al. Efeitos da suplementação de creatina e do treinamento de potência sobre a performance e a massa corporal magra de ratos. Revista Brasileira de Medicina do Esporte. Vol. 13. 2007. p. 297-302.

18-Garrido, R.G.; et al. Suplementação de Creatina por Praticantes de Musculação de Vitória da Conquista/BA. Revista Brasileira de Ciência e Movimento. Vol. 16. Num. 4. 2008.

19-Goston, J.L.; Mendes, L.L. Perfil nutricional de praticantes de corrida de rua de um clube esportivo da cidade de Belo Horizonte-MG, Brasil. Revista Brasileira de Medicina do Esporte. Vol. 17. 2011. p. 13-17.

20-Gualano, B.; et al. Efeitos da suplementação de creatina sobre força e hipertrofia muscular: atualizações. Revista Brasileira de Medicina do Esporte. Vol. 16. 2010. p. 219-223.

21-Gualano, B.; Artioli, G.G.; Lancha Junior, A.H. Suplementação de creatina e metabolismo de glicose: efeitos terapêuticos ou adversos? Revista Brasileira de Medicina do Esporte. Vol. 14. 2008. p. 478-478.

22-Harris, R.C.; Söderlund, K.; Hultman, E. Elevation of creatine in resting and exercised muscle of normal subjects by creatine supplementation. Clinical Science. Vol. 83. Num. 3. 1992. p. 367-374.

23-Haun, C.T.; et al. Pre-training Skeletal Muscle Fiber Size and Predominant Fiber Type Best Predict Hypertrophic Responses to 6 Weeks of Resistance Training in Previously Trained Young Men. Frontiers in Physiology. Vol. 10. 26 mar. 2019.

24-Henry, D.; et al. Clustering Methods with Qualitative Data: a Mixed-Methods Approach for Prevention Research with Small Samples. Prevention Science. Vol. 16. Num. 7. 2015. p. 1007-1016.

25-Hultman, E.; et al. Muscle creatine loading in men. Journal of Applied Physiology. Vol. 81. Num. 1. 1996. p. 232-237.

26-Jäger, R.; et al. International Society of Sports Nutrition Position Stand: protein and exercise. Journal of the International Society of Sports Nutrition. Vol. 14. 2017. p. 20.

27-Jagim, A.R.; et al. Safety of Creatine Supplementation in Active Adolescents and Youth: A Brief Review. Frontiers in Nutrition. Vol. 5. 2018. p. 115.

28-Kreider, R.B.; et al. ISSN exercise & sport nutrition review: research & recommendations. Journal of the International Society of Sports Nutrition. Vol. 7. 2010. p. 7.

29-Kreider, R.B.; et al. International Society of Sports Nutrition position stand: safety and efficacy of creatine supplementation in exercise, sport, and medicine. Journal of the International Society of Sports Nutrition. Vol. 14, 2017. p. 18.

30-Kreider, R.; Jung, Y. Creatine supplementation in exercise, sport, and medicine. The Journal of Exercise Nutrition and Biochemistry. Vol. 6. 2011. p. 53-69.

31-Kumar, S.; Chong, I. Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States. International Journal of Environmental Research and Public Health. Vol. 15. Num. 12. 2018. p. 2907.

32-Liu, H.; Cocea, M. Semi-random partitioning of data into training and test sets in granular computing context. Granular Computing. Vol. 2. Num. 4. 2017. p. 357-386.

33-Lollo, P.C.B. Perfil dos alunos das academias de ginástica de Campinas, SP. Lecturas: Educación física y deportes. Num. 76. 2004. p. 15.

34-Martins, L.I.P.A. Hábitos de consumo de suplementos desportivos com aminoácidos de cadeia ramificada (BCAA). master Thesis. Escola Superior de Tecnologia da Saúde. 2022.

35-Naranjo, S.I.L.; et al. Analysis of the use of the Python programming language for statistical calculations. Espirales revista multidisciplinaria de invesitgación científica. Vol. 6. Num. 2. 2022. p. 1-13.

36-Peralta, J.; Amancio, O.M.S. A creatina como suplemento ergogênico para atletas. Revista de Nutrição. Vol. 15. 2002. p. 83-93.

37-Raschka, S.; Patterson, J.; Nolet, C. Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. Information. Vol. 11. Num. 4. 2020. p. 193.

38-Reis, E.L.D.; et al. Utilização de recursos ergogênicos e suplementos alimentares por praticantes de musculação em academias. Revista Brasileira de Nutrição Esportiva. São Paulo. Vol. 11. Num. 62. 2017. p. 219-231.

39-Rodriguez, N.R.; et al. Position of the American Dietetic Association, Dietitians of Canada, and the American College of Sports Medicine: Nutrition and athletic performance. Journal of the American Dietetic Association. Vol. 109. Num. 3. 2009. p. 509-527.

40-Sousa, M.V.C. Uso de suplementos nutricionais em desportistas portugueses de alto nível das modalidades de atletismo, natação e trialto. FCAUP. 2008.

41-Temkin, S. M.; et al. Chronic conditions in women: the development of a National Institutes of health framework. BMC Women's Health. Vol. 23. 2023. p. 162.

42-Theodoro, H.; Ricalde, S.R.; Amaro, F.S. Avaliação nutricional e autopercepção corporal de praticantes de musculação em academias de Caxias do Sul-RS. Revista Brasileira de Medicina do Esporte. Vol. 15. 2009. p. 291-294.

43-Varghese, J.; Muntode Gharde, P.A Comprehensive Review on the Impacts of Smoking on the Health of an Individual. Cureus. Vol. 15. 2023. n. 10. 2023. p. e46532.

44-WHO. report on the global tobacco epidemic, 2023: protect people from tobacco smoke. Disponível em: <https://www.who.int/publications/i/item/9789240077164>.

45-Wyss, M.; Kaddurah-Daouk, R. Creatine and creatinine metabolism. Physiological Reviews. Vol. 80. Num. 3. 2000. p. 1107-1213.

4 - CI-ISCE, Instituto Superior de Ciências Educativas do Douro (ISCE Douro), 4560-547 Penafiel, Portugal.
5 - Research Centre for Active Living and Wellbeing (Livewell), Instituto Politécnico de Bragança, Portugal.
6 - Department of Sports Sciences, Polytechnic Institute of Guarda, 6300-559 Guarda, Portugal.
7 - SPRINT, Sport Physical activity and health Research & Inovation Center, Guarda, Portugal
8 - Research Center in Sports Sciences, Health Sciences & Human Development (CIDESD), 5001-801 Vila Real, Portugal.

Corresponding Author:
Pereira AM, Department of Nutrition and Dietetics; Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal. Email: amgpereira@ipb.pt

Recebido para publicação em 15/09/2024
Aceito em 08/10/2024