Dec 12th, 12:00 AM

Data Analytics for Uncovering Fraudulent Behaviour in Elite Sports

Maxx Richard Rahman
Saarland University, m.rahman@iss.uni-saarland.de

Thomas Piper
German Sport University Cologne (DSHS), t.piper@biochem.dshs-koeln.de

Hans Geyer
German Sport University Cologne (DSHS), h.geyer@biochem.dshs-koeln.de

Tristan Equey
World Anti-Doping Agency (WADA), tristan.equey@wada-ama.org

Norbert Baume
World Anti-Doping Agency (WADA), norbert.baume@wada-ama.org

See next page for additional authors

Presenter Information

Maxx Richard Rahman, Thomas Piper, Hans Geyer, Tristan Equey, Norbert Baume, Reid Aikin, and Wolfgang Maass

Data Analytics for Uncovering Fraudulent Behaviour in Elite Sports

Completed Research Paper

Maxx Richard Rahman
Saarland University
Germany
m.rahman@iss.uni-saarland.de

Thomas Piper
German Sport University Cologne
Germany
t.piper@biochem.dshs-koeln.de

Hans Geyer
German Sport University Cologne
Germany
h.geyer@biochem.dshs-koeln.de

Tristan Equey
World Anti-Doping Agency
Canada
tristan.equey@wada-ama.org

Norbert Baume
World Anti-Doping Agency
Canada
norbert.baume@wada-ama.org

Reid Aikin
World Anti-Doping Agency
Canada
reid.aikin@wada-ama.org

Wolfgang Maass
Saarland University
Germany
wolfgang.maass@iss.uni-saarland.de

Abstract

Sports officials around the world are facing societal challenges due to the unfair nature of fraudulent practices performed by unscrupulous athletes. Recently, sample swapping has been raised as a potential practice where some athletes exchange their doped sample with a clean one to evade a positive test. The current detection method for such cases includes laboratory testing like DNA analysis. However, these methods are costly and time-consuming, which goes beyond the budgetary limits of anti-doping organisations. Therefore, there is a need to explore alternative methods to improve decision-making. We presented a data analytical methodology that supports anti-doping decision-makers on the task of athlete disambiguation. Our proposed model helps identify the swapped sample, which outperforms the current state-of-the-art method and different baseline models. The evaluation on real-world sample swapping cases shows promising results that help advance the research on the application of data analytics in the context of anti-doping analysis.

Keywords: Data Analytics, Fraud Detection, Doping, Sample Swapping, Machine Learning, Sports

Introduction

Large sports events, such as the Olympic Games or FIFA World Championship, attract the attention of billions of people. Illegal performance enhancement by herbs and other substances can be rooted back into the Olympic Games of Ancient Greece. In recent times, the case of Lance Armstrong disclosed massive doping in cycling but also unleashed investigations in other sports. Consequently, the World Anti-Doping Agency (WADA) was founded with the goal of identification and prosecution of athletes found guilty of taking illegal substances, called doping. It was agreed that the whole process should be based on scientific methods that should guarantee objective decision-making. Anti-doping analytics mainly use methods inherited from Biology and Biochemistry by analysing urine and blood samples taken from athletes during competition and beyond. The success of machine learning technologies posed the question of applicability for doping analytics, which was investigated by a few studies (Sottas et al., 2006). Decision-making on doping lacks confidence in ground truth but instead uses evidence-based truth qualified by domain experts. Therefore, it is necessary that doping analytical procedures are strictly enforced during taking the sample and processing samples in certified doping laboratories. Any breach in procedure compromises the whole anti-doping framework. During the Olympic Winter Games at Sochi, a subsequent report found that at least two female ice hockey players' samples were swapped with a urine sample containing male DNA, and others were found guilty of tampering with the original samples (McLaren, 2016). Urine swapping is the act of exchanging urine with another individual's or the athlete's stored clean urine to evade a positive test (WADA, 2020). More than 1000 athletes across 30 sports were involved in large-scale sample swapping at Sochi 2014. It was a massive program of cheating and cover-ups that has been running on an unprecedented scale since 2011 and will increase in future events (McLaren, 2016). This simple but new form of doping became a threat for the whole anti-doping decision-making organisation. No known statistical method existed so far, and experts doubted that machine learning could do any better.

As described in WADA's Technical Document TD2021EAAS (WADA, 2021), the testing laboratory follows a standardised procedure to quantify the steroid profile markers. First, the Initial Testing Procedure (ITP) is conducted to estimate the steroid profile of the sample. ITP includes the quantification of the concentration of each biomarker by using Gas Chromatography combined with Mass Spectrometry (GC-MS^n) (Mareck et al., 2008). The testing laboratory updates the report of the analysed sample to the Anti-Doping Administration & Management System (ADAMS) (WADA, 2021). In ADAMS, the Adaptive Model based on the Bayesian approach (Sottas et al., 2006) flags profiles for closer examination if one or several specific biomarkers of doping vary beyond its personalised thresholds. These values of these biomarkers could vary beyond its personalised thresholds due to many reasons e.g., due to the intake of doping substance, intake of medication, high altitude training or even sample swapping. Therefore, the subsequent Confirmation Procedure (CP) is performed, including analytical methods like GC-MS^n validation analysis and further Gas Chromatography/Combustion /Isotope Ratio Mass Spectrometry (GC/C/IRMS) (Becchi et al., 1994) to investigate the reason behind the unusual values of the biomarkers. In the suspicious case of sample swapping, after experts review the DNA analysis is conducted by following the relevant technical document TD2021APMU and TD2021 EAAS (WADA, 2021).

However, there are some challenges associated with the current detection methods. Firstly, DNA analysis is not only a time-consuming procedure but also an expensive method to be conducted on all the suspicious samples. Usually, during large sports events like Olympics Games, there are several thousands of samples collected and analysed during the event. However, several hundreds of them are even flagged as suspicious sample swapping case by the Adaptive Model. Conducting DNA analysis on all these samples requires more resources like time and money, which is beyond the budgetary limits and capabilities of anti-doping organisations and associated laboratories. Secondly, the Adaptive Model is a significant element in the current procedure of finding suspicious samples. The Bayesian approach (Sottas et al., 2006) is used to determine the personalised threshold for each biomarker which is used to compare the new samples. These thresholds correspond to a critical range defined by a given specificity assuming a normal physiological condition. Since these thresholds are calculated from the prior distribution based on the reference population, it is effective in flagging the suspicious sample due to the intake of a doping substance, i.e. steroid doping but not effective in finding the sample swapping. Therefore, a new approach is required for a better interpretation of sample swapping, which is time and cost-effective.

In the recent past, the data-driven approach has shown several promising applications in healthcare as well as forensic science (Sidey-Gibbons and Sidey-Gibbons, 2019; Carriquiry et al., 2019). Therefore, it gives a

motivation to explore the data analytical approach that includes not only statistical methods but also machine learning algorithms. This leads us to ask a question, 'Can machine learning help in anti-doping analysis for sample swapping?' To answer this question, it is first important to understand how data-driven methodology can help to improve the anti-doping analysis. Moreover, decisions in anti-doping analysis require a high level of accountability and transparency. An inaccurate decision can lead to several consequences not only for an athlete but also for the nation. Therefore, explanations for the predictions from machine learning algorithms are thus needed to justify their reliability, which requires greater interpretability. So, the main contribution of this study is to perform data analytical research for sample swapping problem and focus on the interpretability aspect. In this paper, we proposed a methodology consisting of a step-by-step process based on data-driven analysis that helps to solve the above-mentioned problems with the current detection method. We performed the data analytics to find the insights from the data and proposed a model which is not only effective in flagging the sample swapping cases but also helps in visualising the closeness of different samples in the steroid profile of the athlete. In the end, we compared the performance of our model with different baseline models and discussed the results.

Related Work

Detection of sample swapping cases has an analogy with the fraud detection problem in forensic science, where fraud refers to any intentionally deceptive action designed to obtain unlawful gain.

Fraud detection

The fraud can be of various types such as transaction fraud, security fraud, insurance fraud etc. There has been a lot of research work done in this domain using a data-driven approach. Pumsirirat and Yan (2018) used deep learning algorithms like deep Autoencoder and restricted Boltzmann machine to detect anomalous credit card transactions which show deviation from normal patterns. The other algorithms like Convolutional Neural Networks are used by Munir et al. (2019) to analyse fraud detection as an anomaly detection problem. Recently, attention models have received a lot of attention in the fraud detection context (Frabmacher et al., 2021). Moreover, several works also attempt to interpret the decision-making of models for fraud detection using explainable AI approaches (Psychoula et al., 2021). Table 1 shows the overview of some recent research work on fraud detection in different domains focusing on machine learning and interpretability aspects.

Literature	Domain	Labels	Data	Fraud	Statistical Analysis	Machine Learning	Interpretability
Kaiafas et al. (2019)	Telecom	S	RW	Voice calls	x	x	
Nguyen et al. (2019)	Security	U	P	Network		x	x
Massi et al. (2020)	Healthcare	U	RW	Diagnosis	x	x	
Li et al. (2020)	Banking	S	RW	Money Laundering		x	x
Troncoso et al. (2020)	Crime	S	RW	Criminal network	x	x	
Zheng et al. (2021)	Banking	S	P	Credit card		x	x
Farbmacher et al. (2021)	Healthcare	S	P	Insurance		x	x
Psychoula et al. (2021)	Banking	U,S	P	Credit card		x	x
Chang et al. (2022)	Finance	S	P	User info		x	
Óskarsdóttir et al. (2022)	Automobile	S	P	Insurance		x	
Our contribution	**Sports**	**U**	**RW**	**Sample swapping**	**x**	**x**	**x**

Table 1. Literature review on the fraud detection in different domain.

S: supervised; U: unsupervised; P: public dataset; RW: real-world dataset

Fraud detection in anti-doping analytics

Some work has been done in anti-doping analytics using a data-driven approach to uncover fraud activities. Doping activities can be classified into blood doping, steroid doping and sample swapping. Most of the data-driven research done until now mainly focuses on blood doping and steroid doping. For example, Rahman et al. (2022) used different machine learning algorithms to detect the presence of doping substance erythropoietin in athletes' blood samples. Kelly et al. (2019) applied different machine learning algorithms with resampling techniques to find athletes at the highest risk of doping based on their performance data. Montagna and Hopker (2018) used a bayesian approach for the detection of blood doping by using the interindividual performance data. These studies mainly focused on blood doping. On the other hand, literature on steroid doping includes Renterghem et al. (2013), who used Support Vector Machine on the athlete's steroid profile to find how much a profile deviates from the normal population profiles. Wilkes et al. (2018) used machine learning algorithms such as Random Forest and XGBoost to predict abnormalities in steroid profiles. Alladio et al. (2016) used a statistical method like Hotelling's T2 test and Principal Component Analysis to detect anomalous steroid profile. All these works consider reference population data to define a normal profile and use different algorithms to find the deviation of anomalous steroid profile. However, this method cannot be used to detect sample swapping where only the samples collected from the same athlete should be considered for defining the normal profile. Therefore, there is a need to explore data-driven methods for the sample swapping problem. However, these studies show that the data analytical approach, especially machine learning is not a new concept in anti-doping control and has been in place to answer several research questions in anti-doping analysis.

However, there is no study performed so far on addressing the problem of sample swapping using a data-driven approach. The current state-of-the-art method (SoTA) for finding the sample swapping is still the Bayesian method of the Adaptive Model, followed by laboratory testing (Sottas et al., 2006). Once triggered by the Adaptive Model, the confirmation tests like IRMS tests or subsequent DNA analysis of the athlete are performed to verify whether the sample is from the same athlete or is substituted by the athlete (Piper et al., 2021; Thevis et al., 2012; Thevis et al., 2006). However, these laboratory-based approaches are too expensive and cannot be implemented on all the athletes' samples during large athletic events like Olympic Games. Moreover, it is also a time-consuming process due to the fact that conducting each confirmation test requires a significant amount of time and resources. This is the reason why in most cases, unscrupulous athletes are caught after several months of the athletic event. This shows why there is a need to explore a new and more efficient method in this direction.

Our model provides a solution to this problem by creating a profile of an athlete based on their samples. In this profile, we can visualise the relatedness of all the samples of the athlete and track the changes when a new sample is added to find the sample swapping case. This method is very cost-effective and detects sample swapping in real-time. Therefore, our model can help the decision makers to flag the sample swapping cases during the athletic event and explain their decisions. In this way, it will reduce the number of laboratory-based testing needed and hence, cost and time beneficial.

Methodology

The proposed methodology used in conducting our analysis is based on the suggestion by Shmueli and Koppius (2011). It consists of a set of components forming a pipeline. Statistical and machine learning analysis are the main form of study in this research. In this section, we give a general overview of the pipeline and explain each component in detail. Fig. 1. shows the schematic representation of our pipeline and its components.

Goal Definition

The goal of this study is to develop a model that can detect the sample swapping activity performed by the athlete. In other words, a model that can tell whether the steroid profile of the urine sample is from the same athlete who gave the sample or not. In addition, the model should trigger whether a new steroid profile

matches with the steroid profile of previous samples of the athlete collected over time. So, a visualisation tool or a quantification measure is needed that can show the relatedness among the steroid profiles of the same athlete. Therefore, a comprehensive analysis of steroid profiles has to be conducted to understand the underlying principles of different biomarkers.

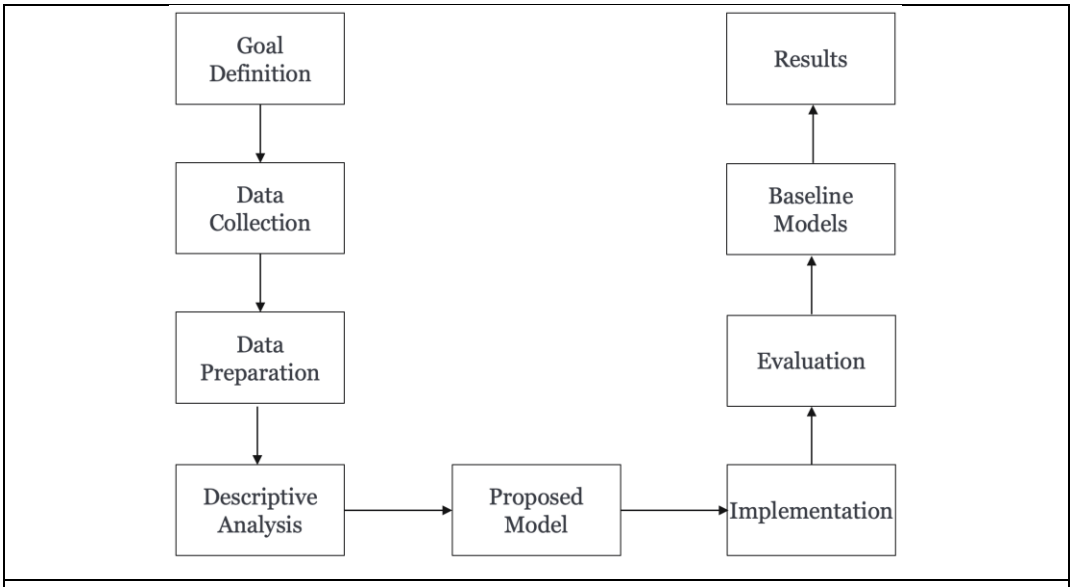

Figure 1. Graphical representation of the methodology and its components.

Data Collection

Sport federations and anti-doping agencies across the world conduct doping tests throughout the year at various national and international athletic events. This presents large scale historical data for each individual athlete. The data is extracted from the ADAMS database, that consists of the real-world athlete data collected from 1 September 2018 until 31 March 2021. This data contains 254,478 urine samples corresponding to 65,039 athletes where each athlete could have between 2-20 samples in their profile. Table 2 shows the summary of the number of samples belonging to male and female athletes. For each athlete, we extracted only the raw steroid profile values, the gender, the competition type whether tested during competition (INC) or out of competition (OOC), the specific gravity of the sample (SG) and an anonymised athlete ID into an anonymised dataset in accordance with the WADA International Standard for the Protection of Privacy and Personal Information (ISPPPI) (WADA, 2021).

	Profiles	Samples
Male athletes	52,152	166,237
Female athletes	12,887	88,241
Total	65,039	254,478

Table 2. No. of samples and the profiles belonging to male and female athletes in the data.

The steroid profile of the urine samples consists of a set of biomarkers called steroid parameters that show significant changes in the administration of steroids. These parameters are Androsterone (A), Etiocholanolone (Etio), Epitestosterone (E), Testosterone (T), 5α-androstane-3α,17β-diol (5aAdiol), 5β-androstane-3α,17β-diol (5bAdiol) and their ratios T/E, A/Etio, A/T, 5aAdiol/5bAdiol, 5aAdiol/E as described in TD2021EAAS (WADA, 2021).

Data Preparation

Missing values

In the data, we found that there are some samples contained missing values marked with '0' for all the parameters, which could mainly be due to some error in extracting the data from the ADAMS database. However, it could also be possible that the test result was not reported/updated in the ADAMS due to analytical issues by the testing laboratory. Since we are interested in finding the similarity among the samples in the longitudinal profile of the athlete, imputing the values from the complete dataset (i.e., using the samples from other profiles/athletes) will not be a possible solution as it introduces bias in the profile. Therefore, we removed all the samples with missing values. Fig. 2 shows the data distribution of raw data and data statistics after removing the samples containing the missing values for male and female athletes.

Reference ranges, LOQ and LOD

The values for each steroid parameter should lie within a certain range, but we observed some exclusively high values of certain parameters in few samples. Therefore, we compared these values with the maximum value of that parameter from Renterghem et al. (2010) and removed all the samples containing off-values.

Limit of Quantification (LOQ) refers to when the laboratory cannot quantify the concentration of the steroid parameter by GC-MS^n, and therefore, is reported as '-1', whereas Limit of Detection (LOD) refers to when the chromatography peak signal of the parameter cannot be detected (i.e., is below the detection capability of the assay) and reported as '-2'. We replaced these values with the lowest concentration values that can be measured with uncertainty not greater than 30%, as mentioned in Technical Document TD2021EAAS (WADA, 2021).

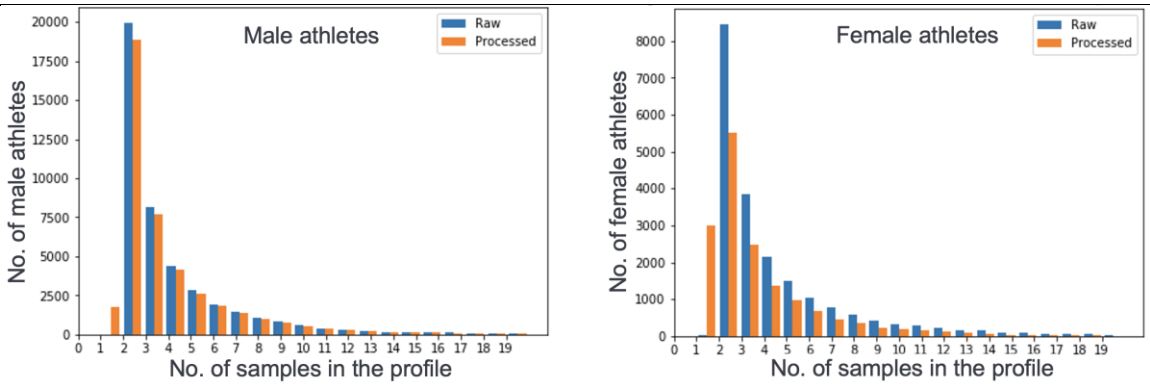

Figure 2. Distribution of no. of samples in the profile per athlete for male (left) and female (right) athletes. Raw data (blue) represents the original dataset and processed data (orange) represents the data after removing the samples containing missing values.

Correction due to urinary concentration

Not all the collected samples have the same concentration since some are more diluted than others. To compare the measured concentrations between different samples, the urinary concentrations need to be normalised using the urinary density. The concentration value of Testosterone parameter of all the samples was corrected to a specific gravity of 1.020 as given by TD2021DL (WADA, 2021):

$$T = \frac{1.020 - 1}{SG - 1} \times T_{raw}$$

where T_{raw} represents the concentration value before the correction is applied and SG represents the specific gravity of the sample. Similarly, the correction for A, Etio, E, 5aAdiol and 5bAdiol was also applied. The steroid ratios are unaffected by the urinary specific gravity.

Descriptive Analysis

The distribution of all the steroid parameters of the samples was tested by using the 2-sample Kolmogorov-Smirnov test (K-S test) (Dimitrova et al., 2020). It is a standard test for deciding whether the two distributions are consistent with each other. We performed the K-S test between the distribution of samples collected during the competition and out of the competition for both male and female athletes. The *p*-value for each steroid parameter shows that there is a significant influence on the steroid profile of the samples due to the testing during the competition. In a recent laboratory study, Piper et al. (2021) also found that there is a confounding factor due to the physical and mental stress on athletes, which causes a significant amount of elevated values in the profile. Our statistical results show a similar change and thus are consistent with their findings.

Fig. 3 shows the statistical distribution of testosterone parameter for male and female athletes. We can observe that the testosterone values are more sparse for male athletes than female athletes. Due to which, there is less inter-individual variance among female samples. Moreover, we found that there is a linear correlation between the steroid parameters. The parameters A, Etio, E, T, 5aAdiol and 5bAdiol represent the concentration values of different steroids in the steroid metabolism pathway. The concentration value of one steroid affects the other parameters, as described in Piper et al., (2021). In addition, there is a linear relationship between these 6 parameters and 5 ratio parameters. For example, T is directly proportional to T/E parameter. Therefore, we observed a collinearity among the steroid parameters. We have statistically described the distributions of the steroid parameters using mean, median, first (IQ1 = 0.25), and third (IQ3 = 0.75) quartile. Fig. 4 and Fig. 5 show the detailed descriptive statistics of the steroid parameters of male and female athletes, respectively.

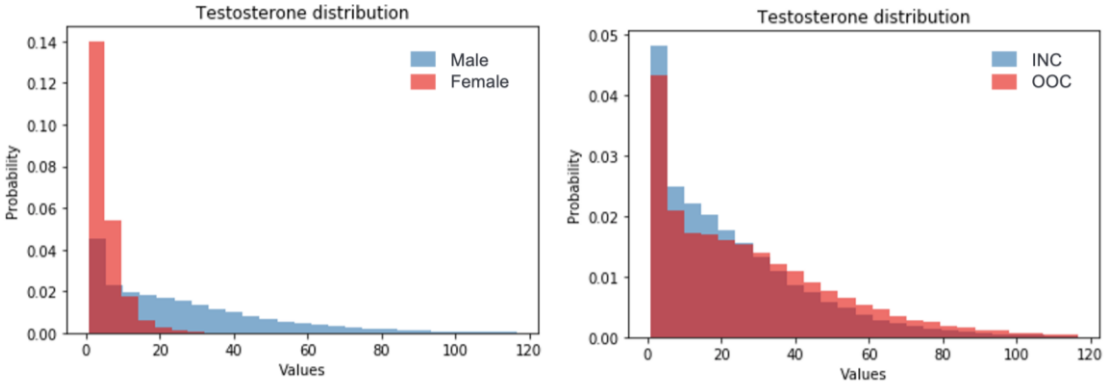

Figure 3. Distribution of Testosterone parameter between male and female athletes (left) and between INC and OOC samples of male athletes (right).

Proposed Model

Data visualisation is an important concept in the data-driven approach (Vellido et al., 2011). It helps to explore data structure, detect outliers, identify trends/patterns or even interpret the result to gain information. Therefore, it is important to visualise the steroid samples in either two- or three-dimensional space. However, since the steroid profile consists of 11 parameters representing different biomarkers, the elevated values of any of these biomarkers can significantly impact the other biomarkers. Therefore, the steroid profile should be visualised in a space that takes into account all the parameters at the same time. For example, let us consider a three-dimensional space spanned by any three arbitrary chosen parameters. We will require a total of 165 different spaces to completely visualise a steroid profile to view all the aspects. Fig. 6 shows an example of the longitudinal steroid profile of an athlete with a testing sample in 8 different spaces (out of 165 spaces) spanned by three different arbitrary chosen steroid parameters. Based on these plots, it is difficult to state whether the testing sample belongs to the same athlete profile or from another athlete since there is no evidence of which space should be considered for decision making. Therefore, there is a need to find a visualisation aid that incorporates the behaviour of all the parameters together.

	INC (n=73053)						OOC (n=88096)						p-value
Parameter	mean±*std.*	min	IQ1	median	IQ3	max	mean±*std.*	min	IQ1	median	IQ3	max	
A	3084±1666	17.3	1925	2756	3882	37205	2604±1390	19	1634	2345	3274	28690	1.2e-14
Etio	2038±1101	40.5	1265	1810	2556	17500	2062±1125	31.1	1275	1833	2580	24326	3.8e-99
E	26.4±19.8	0.45	12.8	21.3	34.1	477	30.8±22.2	0.6	15.1	25.3	40.6	562	0.0
T	28.9±23.7	0.2	10.9	25.0	40.4	399	31.4±24.8	0.3	12.3	27.9	44.0	334	0.0
5aAdiol	58.1±40.2	2.1	32.2	49.1	73.1	1050	58.0±38.2	1.1	32.2	49.3	73.7	1523	3.8e-99
5bAdiol	131±116	2.5	50	97.9	174	2157	152±132	1.1	58.1	116	204	2433	0.0
T/E	1.4±1.2	0.0	0.6	1.1	1.9	61.5	1.3±1.2	0.0	0.6	1.1	1.8	46.7	1.7e-60
A/Etio	1.7±0.8	0.0	1.1	1.6	2.0	110	1.4±0.7	0.0	0.9	1.3	1.7	76.6	0.0
A/T	279±428	0.3	71.2	111	227	18602	216±319	1.0	56.0	84.4	164	10470	0.0
5aAdiol/5bAdiol	0.7±0.5	0.0	0.3	0.5	0.9	14.9	0.6±0.5	0.0	0.3	0.4	0.8	13.8	9.3e-99
5aAdiol/E	3.0±2.7	0.0	1.4	2.3	3.6	249	2.5±2.3	0.0	1.2	1.9	3.1	172	0.0

Figure 4. Descriptive statistics of steroid parameters for in-competition and out-competition samples for male athletes including the *p*-value from the K-S test.

	INC (n=34320)						OOC (n=49940)						p-value
Parameter	mean±*std.*	min	IQ1	median	IQ3	max	mean±*std.*	min	IQ1	median	IQ3	max	
A	2335±1526	7.1	1305	1983	2956	27102	1787±1200	9.0	969	1509	2282	26054	1.9e-53
Etio	2178±1311	28.7	1290	1900	2747	34159	1973±1204	9.2	1130	1705	2521	24391	1.5e-16
E	9.4±7.5	0.5	4.4	7.4	12.1	210	8.3±7.0	0.3	3.6	6.3	11.0	174	5.3e-26
T	7.3±7.5	0.1	2.5	5.6	9.9	254	4.6±4.4	0.1	1.8	3.6	6.0	123	3.8e-99
5aAdiol	26.1±26.2	0.2	12.9	20.0	31.4	1900	22.1±20.0	0.1	10.6	17.0	27.2	717	3.1e-11
5bAdiol	58.5±63.5	0.2	20.4	38.3	73.0	1600	66.0±70.6	0.1	20.0	42.0	86.2	1133	1.1e-99
T/E	1.0±1.0	0.01	0.34	0.81	1.36	38.9	0.8±0.9	0.0	0.3	0.6	1.0	125	6.5e-99
A/Etio	1.2±0.7	0.0	0.8	1.1	1.4	50.1	1.0±0.5	0.1	0.6	0.9	1.2	27.3	0.0
A/T	694±1043	0.1	209	346	679	22150	698±954	1.4	252	407	719	21718	4.2e-99
5aAdiol/5bAdiol	0.7±0.5	0.0	0.3	0.6	0.9	25.0	0.6±0.5	0.0	0.2	0.4	0.8	9.6	1.7e-99
5aAdiol/E	3.7±3.8	0.0	1.7	2.8	4.6	161	3.9±4.4	0.0	1.6	2.8	4.8	472	7.8e-11

Figure 5. Descriptive statistics of steroid parameters for in-competition and out-competition samples for female athletes including the *p*-value from the K-S test.

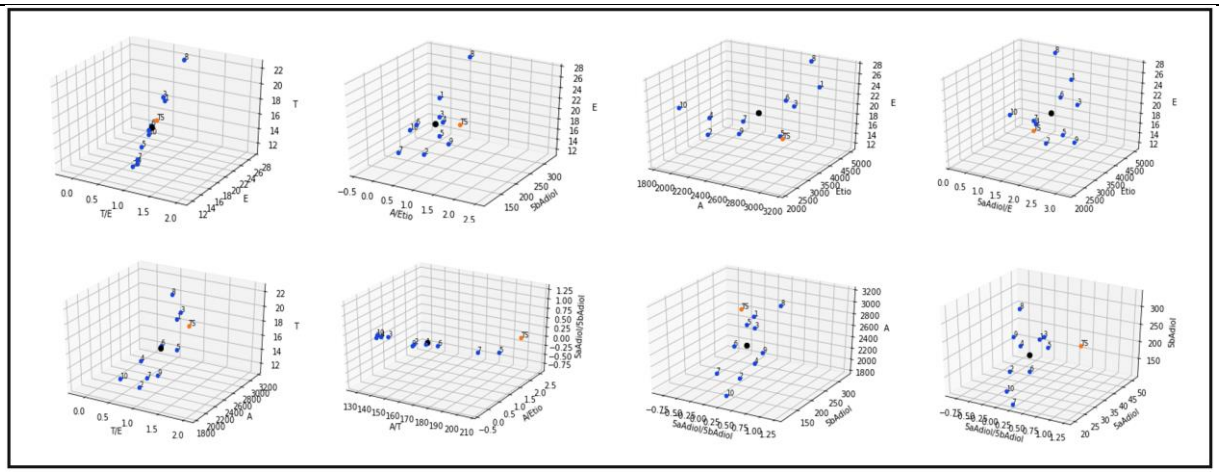

Figure 6. Longitudinal steroid profile of the same athlete in 8 different three-dimensional space span by any 3 arbitrary chosen steroid parameters.

We propose a model called Digital Athlete Passport (DAP), which is an effective approach for understanding the relatedness among the steroid profiles and provides a complete visualisation concept for steroid profiles. DAP consists of Principal Component Analysis and the concept of centroid to illustrate the similarity between the steroid samples of the athlete. We used these techniques instead of other methods because of the following reasons: 1) Since there is a linear relationship between the steroid parameters, PCA helps to reduce correlated parameters to a smaller set of mutually-independent components that explain a large percentage of the covariance in the original steroid parameter space. Other dimension reduction algorithms, like autoencoders, require a large training dataset. Since each athlete's longitudinal profile consists of 2-20 samples, it is not a good choice. 2) Moreover, PCA with the centroid approach also helps to solve the visualisation problem by mapping the steroid sample from a multi-dimensional space to three-dimensional space and tracking the changes in the overall profile of the athlete when a new sample is added.

Principal Component Analysis

Principal Component Analysis (PCA) is an unsupervised learning technique (Lever et al., 2017) that projects the data into a new space spanned by a set of basis vectors such that the maximum amount of information is preserved in a lower number of basis vectors of the new space. The data is projected on these basis vectors called principal components, which are orthogonal unit vectors that maximise the variance in the data.

The weights of each principal component represented by $w(k)$ is calculated by the following expression:

$$w(k) = argmax\left\{\frac{w^T X(k)^T X(k) w}{w^T w}\right\}$$

where $k = \{1,2,3\}$ and X refers to the data. The transformed data X' can be obtained by:

$$X' \in x_i'(k) = x_i . w(k)$$

The data is transformed in such a way that it contains the maximum variance in the first component, the second maximum variance in the second component and so on. In our model, we used PCA to transform the steroid profile (consists of 11 parameters) into a set of 3 principal components.

Centroid

The concept of Centroid/Center-of-Mass (CoM) is common in classical mechanics (Kleppner et al., 1973) which has a useful application in many domain. It refers to a unique point in the space where the weighted relative position of the distributed points sums to zero. This means if we have different points spanned in the space, CoM represents the approximate center of all these points and can be calculated by the following expression.

$$x'_{CoM}(k) = \frac{1}{N}\sum_{i=1}^{N} x_i'(k) \qquad k = 1, 2, 3$$

where $x'_{CoM}(k)$ represents the centroid of all the transformed samples in the longitudinal profile with k representing the three principal components of the transformed sample and N represents the number of steroid samples in the longitudinal profile.

Whenever a new steroid sample of the athlete is added, it is important to measure the relatedness of this sample with respect to the previous samples. We can solve this problem by tracking the position of the CoM. If the position of the new steroid sample is distant from the previous samples, then it will cause a large deviation in the position of the CoM or vice versa. So, the variation in the position of CoM could be a useful measure to monitor the consistency among steroid samples in the longitudinal profile.

Implementation

In the Digital Athlete Passport, we take the longitudinal profile of the athlete which consists of all the samples and the testing sample (TS), which needs to be checked for sample swapping:

- **Step 1:** Since we are interested in visualising the longitudinal profile of the athlete in three-dimensional space, we need to calculate three principal components of each of the steroid samples. Therefore, we

require at least three samples in the profile. So in this analysis we considered the longitudinal profiles which consists of atleast 3 samples. The PCA is performed on the first three steroid samples of the athlete's longitudinal profile after randomising the order of the samples to remove any kind of bias. The calculated weights for each principal component are then used to transform the next steroid sample in the profile.

- **Step 2:** The CoM point is calculated based on the transformed samples by the principal components. This process is iterated until all the samples of the profile are considered, including the testing sample. This excludes the collinearity between the original parameters and hence provide better values.
- **Step 3:** We plot the three components of all the transformed samples in a three-dimensional space since it captures most of the variance of the athlete's longitudinal profile.

Fig. 7 shows a randomly selected athlete's longitudinal profile in three arbitrary chosen steroid parameter space (SPS) and the transformed samples (after applying the DAP algorithm) in the principal component space (PCS). CoM point (in black) shows the arithmetic centre of all the transformed samples. The position of the testing sample (in orange) with respect to the CoM and other samples shows the likelihood whether the testing sample belongs to the same athlete profile. In this case, we can observe in PCS that the testing sample does not belong to the same athlete profile, which was not evident in SPS. The pseudocode of the detailed DAP algorithm is shown in Fig. 8.

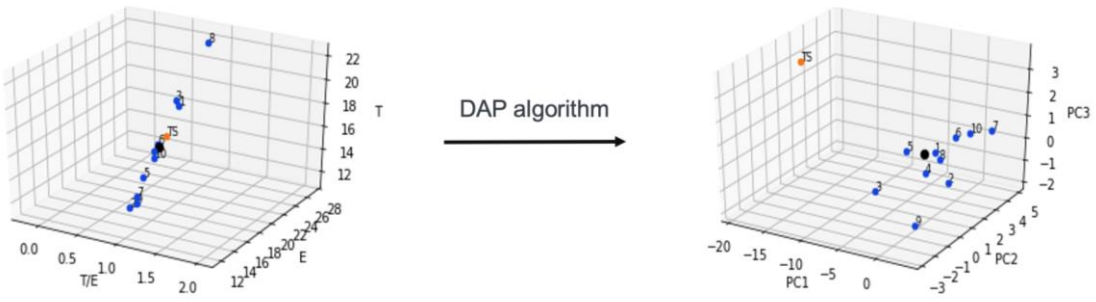

Figure 7. Athlete's longitudinal profile in steroid parameter (SPS) (left) and the transformed profile in principal component space (PCS) (right) after applying the DAP algorithm.

Consecutive distance

It is important to understand the change in the characteristics of the athlete's longitudinal profile when a new steroid sample is collected and added to the profile. In DAP, this can be done by tracking the position of the CoM whenever a new steroid sample is added. The intuition behind this is that the more alike the new sample is from the previous samples, the more the CoM will deviate. Therefore, we also calculate the consecutive distances between the position of CoM on the addition of each sample, as shown in Fig. 8b. We use Euclidean geometry for the distance computation using the following expression:

$$d_i = \sqrt{\sum_{k=1}^{3} \left(x_{CoM}^{i}(k) - x_{CoM}^{i-1}(k)\right)^2}$$

where d_i represents the distance shifted by CoM when the i^{th} sample is added to the profile, and k represents the three components of the CoM in PCS.

Cumulative distance

We also calculate the total distance deviated by CoM after all the samples are added to the profile, as shown in Fig. 8c. This helps to keep track of how much the characteristic of the profile gets impacted on the

addition of a new sample. We observed that as soon as we have more samples in the longitudinal profile, there is less impact on the position of CoM and d_i starts decreasing unless there is a suspicious sample from a different athlete. In such a case, we can observe a sudden spike in the plot. The cumulative distance is calculated by the following expression:

$$D_i = \sum_{m=3}^{i} d_m$$

where D_i represents the cumulative distance until i^{th} sample is added. Since we start with the first three samples for calculating the CoM, therefore m starts with 3. This is because we need the three components of the transformed sample for the DAP algorithm.

Algorithm Digital Athlete Passport algorithm

Require: $N_{sample} \geq 3$
  $randomize()$
  $X \leftarrow x_1, x_2, x_3$
  for $k \leq N_{parameter}$ **do**
    $w(k) \leftarrow PCA(X)$ ▷ calculate weights
    $k \leftarrow k + 1$
  end for
  while $i \neq N_{sample}$ **do**
    for $k \leq 3$ **do**
      $x'_i(k) \leftarrow x_i(k).w(k)$
      $k \leftarrow k + 1$
    end for
    if $i \geq 3$ **then**
      $x_{CoM} \leftarrow \frac{1}{i} \sum_{m=1}^{i} x'_m$ ▷ center-of-mass
      $d \leftarrow |x_{CoM} - x_{CoMi-1}|$ ▷ consecutive distance
      $D \leftarrow D + d$ ▷ cumulative distance
      $i \leftarrow i + 1$
    else
      $i \leftarrow i + 1$
    end if
  end while

Figure 8. Pseudocode of Digital Athlete Passport algorithm.

Contribution of each steroid parameter

It is important to understand the contribution of steroid parameters to each principal component. This helps to determine the importance of the steroid parameter for decision-making. Therefore, we calculated the feature importance of each steroid parameter for the athlete's longitudinal profile in PCS. The importance of each feature can be reflected by the magnitude of the corresponding absolute values in the eigenvectors of $X^T X$, i.e. the larger the absolute values, the more feature contributes to that principal component. In DAP, we calculate it for each longitudinal profile separately, as shown in Fig. 8d.

Variance captured by each component

Each principal component captures a certain amount of variance in the longitudinal profile data. We calculated the proportion of total variance captured by the three principal components, as shown in Fig. 8e.

Evaluation

Our model provides a visualisation aid to understand the similarity of the samples. Thus, the primary way to evaluate whether a sample belongs to the same athlete can be done by domain experts (human evaluation) after the application of the DAP algorithm on the longitudinal profile of the athlete.

However, we also proposed an additional evaluation metric to quantify the sample swapping using DAP algorithm. Let us consider the distance between the CoM and each sample of the longitudinal profile (in DAP) to be $d_1, d_2,$ and so on, and the distance between CoM and testing sample be d_{TS}. We calculate the mean (μ_d) and standard deviation (σ_d) of all the distances. The idea is to compare d_{TS} with the distribution of d to classify the testing sample as an outlier by the following expression:

$$Decision = \begin{cases} Swap, & d_{TS} > \mu_d + 3\sigma_d \\ Not\ swap, & d_{TS} \leq \mu_d + 3\sigma_d \end{cases}$$

The decision rule of $3\sigma_d$ is chosen after performing the sensitivity analysis on the training dataset. We divided the complete dataset into training set (80%) and testing set (20%) as shown in Table 3. In both the dataset, we randomly selected 50% of profiles and manually swapped the last sample with a sample from a different athlete and labelled them as swapped profiles (class 1). The other 50% of the profiles were labelled as clean profiles (class 0). This is performed to create a scenario of swapped and clean case for the classification. We performed the sensitivity analysis on the decision rule to understand the impact on different evaluation metrics. Table 4 shows the performance results of our model on training dataset when different values like $1\sigma_d, 1.5\sigma_d, 2\sigma_d, 3\sigma_d$ and $4\sigma_d$ are chosen for the decision rule.

	Male		Female	
	Profiles	Samples	Profiles	Samples
Training	33,618	128,807	12,572	67,498
Testing	8,405	32,342	3,144	16,762
Total	42,023	161,149	15,716	84,260

Table 3. Data statistics of training and testing set.

	Metrics	$\mu_d + 1\sigma_d$	$\mu_d + 1.5\sigma_d$	$\mu_d + 2\sigma_d$	$\mu_d + 3\sigma_d$	$\mu_d + 4\sigma_d$
Male athletes	Accuracy	0.78	0.78	0.79	0.83	0.75
	Sensitivity	0.82	0.75	0.72	0.72	0.56
	Specificity	0.75	0.87	0.85	0.89	0.94
Female athletes	Accuracy	0.72	0.73	0.72	0.74	0.67
	Sensitivity	0.71	0.63	0.62	0.61	0.40
	Specificity	0.74	0.81	0.84	0.88	0.93

Table 4. Sensitivity analysis on the decision rule showing different evaluation metrics.

For example, we created a scenario of a sample swapping case where a longitudinal profile of an athlete is arbitrarily chosen, and a testing sample (selected from another athlete's profile) is added to the longitudinal profile. The DAP algorithm is applied on this longitudinal profile, and Fig. 9 shows the complete result from the DAP algorithm, including the plots showing the consecutive distance, cumulative distance, contribution from each steroid parameter and variance covered by each principal component. The 3D plot shows that the position of the testing sample is far from the CoM and other samples of the athlete. Thus, the expert can easily understand that the testing sample is suspicious. Moreover, the sudden spike in the consecutive distance (and cumulative) covered by CoM when the testing sample is added triggers that the sample does not belong to this athlete's profile. The values for the proposed evaluation metric $d_{TS} = 39.3,\ \mu_d = 8.8$ and $\sigma_d = 4.5$ also suggest that it is a swapped case. In practice, the test sample usually appear in the end i.e., the last sample of the profile. In this example, we showed how our model works independent of the position of the testing sample i.e., the sample order does not matter.

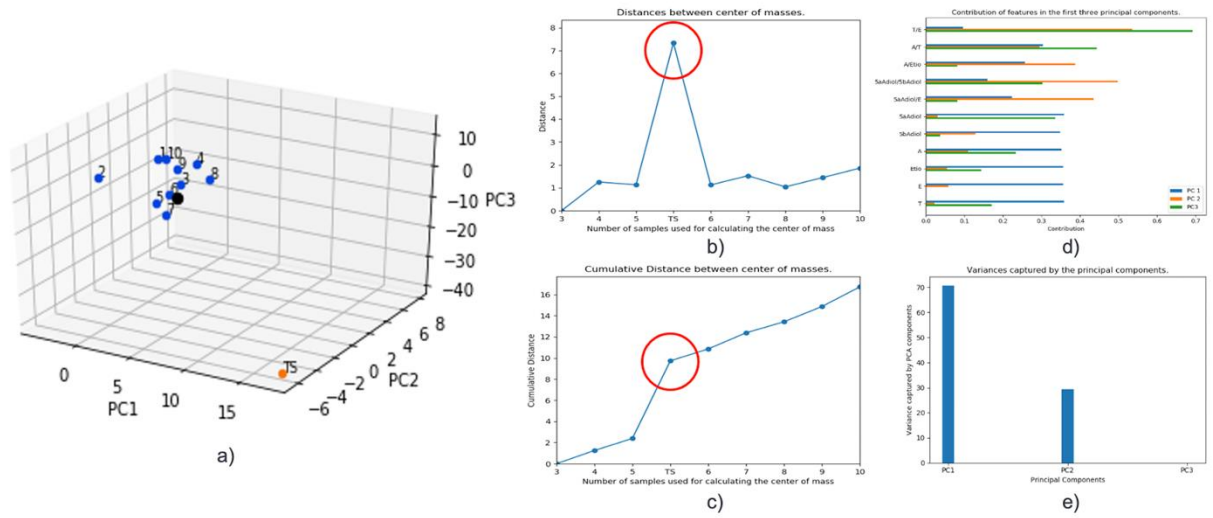

Figure 9. Fully-functional view of the Digital Athlete Passport on arbitrary chosen athlete's longitudinal profile. a) transformed samples in PCS, b) consecutive distance, c) cumulative distance, d) contribution of each steroid parameter and e) proportion of variance captured by each component.

Baseline Models

We selected a set of baseline models which are used to compare the performance of our proposed model. These models are trained and optimised on the training dataset. The models include Logistic Regression (LR) (Peng et al., 2002), Support Vector Machine (SVM) (Hearst et al., 1998), Random Forest (RF) (Breiman, 2001), Gradient Boosting (XGB) (Chen and Guestrin, 2016) and Bayesian Method of Adaptive Model (SoTA) (Sottas et al., 2006). Table 5 shows the different hyperparameters values selected to train each model. These values are considered after performing the optimisation step.

Model	Parameter value
Logistic Regression (LR)	*max iter* = 100 *penalty* = 12
Support Vector Machine (SVM)	*degree* = 3 *kernel* = rbf *max iter* = -1
Random Forest (RF)	*min samples split* = 2 *n estimator* = 100 *bootstrap* = True *criterion* = gini
Gradient Boosting (XGB)	*max depth* = 5 *objective* = binary logistic *n estimators* = 10 *alpha* = 10 *learning rate* = 0.1
Bayesian Method (SoTA)	*specificity threshold* = 99.9%

Table 5. Hyperparameter values of different baseline models used for training.

Results

We performed the evaluation of our model in two phases, first on the longitudinal profiles of the athletes from testing dataset, which we manually swapped to create a sample swapping scenario and the second on the real-world swapping cases, which were confirmed by subsequent DNA analysis.

Evaluation phase I

In the first phase, we selected all the longitudinal profiles in the testing data. For half of these profiles, we chose the testing sample from the same athlete's profile, and for the other half, we chose the testing sample from another athlete's profile. In this way, we manually created a scenario of a sample swapping case where some of the longitudinal profiles consist of testing samples from different athletes. Finally, we applied the DAP algorithm as well as all the trained baseline models. The results are summarised in Table 6.

The result shows that our proposed model could be able to differentiate the swapped profiles from the clean athlete's profiles based on our proposed decision rule. Our model achieves an accuracy of 81% on the longitudinal profiles of male athletes. However, we observed slightly less performance on the profiles of the female athletes. As explained in Fig. 3, this is because the female athletes' samples have less inter-individual variance than the male athletes. The results show that the ensemble method like gradient boosting algorithm shows comparable performance to our proposed model in terms of accuracy and specificity. Since the prevalence of sample swapping cases in very less (<1%) in real-life scenario, high specificity values are important to minimize false positive cases (cost factor). Overall, the proposed model shows better performance than the current state-of-the-art method (Bayesian approach) as well as all the baseline models in terms of different evaluation metrics.

	Metrics	LR	SVM	RF	XGB	SoTA	DAP
Male athletes	Accuracy	0.75	0.73	0.78	0.80	0.76	0.81
	Sensitivity	0.00	0.38	0.25	0.61	0.73	0.75
	Specificity	0.89	0.88	0.92	0.93	0.82	0.92
Female athletes	Accuracy	0.68	0.71	0.75	0.76	0.71	0.77
	Sensitivity	0.00	0.26	0.08	0.48	0.38	0.61
	Specificity	0.84	0.82	0.88	0.90	0.85	0.89

Table 6. Results of our proposed model compared to different baseline models and SoTA method.

Evaluation phase II

In the second evaluation, we validated our model on two real sample swapping cases, confirmed by subsequent DNA analysis by one of the WADA's accredited laboratories. These longitudinal profiles contain more than one samples which were not from the same athlete. The DAP algorithm triggered both the longitudinal profiles successfully and detected them as a sample swapping case, as shown in Table 7. In addition, our model was also applied on the longitudinal profiles that were confirmed positive steroid doping cases (i.e. administration of exogenous steroids) by the laboratory. Again, the model could able to flag these longitudinal profiles as suspicious profiles. The reason we have only two confirmed sample swapping cases for the evaluation shows that, in reality, we have a very less prevalence of the sample swapping cases. Therefore, it is an important and difficult task to identify such cases in anti-doping analysis. These evaluations conclude that our method shows promising results and could potentially improve the current method of detecting sample swapping cases.

Cases	Number of profiles confirmed by laboratory	% of profiles flagged by SoTA method	% of profiles flagged by our model
Sample swapping	2	100%	100%
Steroid doping	5	100%	100%
Normal profiles	23	78%	82%

Table 7. Results of our proposed model on DNA proven profiles compared to SoTA method.

Conclusion and Outlook

The objective of this analysis was to address a research question on whether a data-driven approach can contribute to identifying fraudulent behaviour in sports especially manipulation of samples by the athletes to evade positive doping results. In other words, can machine learning be a helpful approach to identify sample swapping and hence improve the current decision making by solving the problems associated with the current state-of-the-art methods? Indeed, sample swapping has not often been addressed in the design of a data-driven approach. However, several studies have discussed the application of machine learning especially supervised learning algorithms in anti-doping analysis like blood/steroid doping. These studies are often conducted with the help of the data gathered in the clinical trials on individual populations. In our study, we analysed the anonymized real-world data of elite athletes to understand the problem of sample swapping. We proposed a step-by-step process from collecting data to finding the insights of the data to conduct our analysis.

We proposed the model which provides a visualisation aid for the longitudinal steroid profile of the athletes for finding suspicious samples. The model is based on an unsupervised learning algorithm that does not depend on the prior distribution of the steroid samples of the athlete unlike the Bayesian approach. Since the elevated values in different biomarkers give an indication of the suspicious sample, the model transforms the steroid markers into a space where most of the variance is covered and also helps to track the change of the overall profile of the athlete when a new sample is added. Therefore, experts can easily analyse the longitudinal profile of the athlete in DAP and can decide whether to conduct further investigations on the athlete. The application of the proposed model on the steroid profiles of the athletes offers better decision making compared to the SoTA as well as it gives time and cost benefits by a significant amount. Furthermore, the DAP not only triggers the suspicious sample but also helps to explain why the sample is triggered as a suspicious sample. This shows a potential application of DAP as an extension to the current Adaptive Model used by WADA. Therefore, it gives a possible solution to address the statistical and explainability challenges encountered by using the current approach of steroid passport interpretation.

Limitations and Future Work

In this section, we describe the weakness of our model and the possibility of improving it in future research. 1) DAP model requires at least 3 samples in the longitudinal profile of the athlete to detect sample swapping. Therefore, the model cannot be applied to the longitudinal profile of new athletes who had just started their sports career. In such a scenario, a generative algorithm can be integrated into the model to overcome this problem. 2) Since our model leverages the information about variance of the samples for decision making, there is no distinction between the sample triggered by DAP being an actual sample swapping case or could be even steroid doping case like administration of exogenous steroids. Therefore, the model shows potential for further investigation by extending its application not only to flag the suspicious sample but also to be able to distinguish between sample swapping and steroid doping. Therefore, in future, we will analyse more longitudinal profiles in DAP which were confirmed positive with steroid doping or sample swapping by DNA analysis. 3) Moreover, based on the results of evaluation phase I, it is evident that there is room for improving the decision rule of the model. Therefore, a proper formulation of the decision rule needs to be investigated to further improve the performance of the model. 4) In future, we plan to perform empirical studies with experts to understand how the DAP model can help the decision makers to detect sample swapping during sports events.

References

Alladio, E., Caruso, R., Gerace, E., Amante, E., Salomone, A., and Vincenti, M. 2016. "Application of multivariate statistics to the steroidal module of the athlete biological passport: a proof of concept study," *Analytica chimica acta (922)*, pp. 19-29.

Becchi, M., Aguilera, R., Farizon, Y., Flament, M. M., Casabianca, H., and James, P. 1994. "Gas chromatography/combustion/isotope-ratio mass spectrometry analysis of urinary steroids to detect misuse of testosterone in sport," *Rapid communications in mass spectrometry RCM (8:4)*, pp. 304–308.

Breiman, L. 2001. "Random Forests," *Machine Learning (45)*, pp. 5-32.

Carriquiry, A., Hofmann, H., Tai, X.H., and VanderPlas, S. 2019. "Machine learning in forensic applications," *Significance (16),* pp. 29-35.

Chang, J.-W., Yen, N., and Hung, J. 2022. "Design of a NLP-empowered finance fraud awareness model: the anti-fraud chatbot for fraud detection and fraud classification as an instance," *Journal of Ambient Intelligence and Humanized Computing*.

Chen, T. and Guestrin, C. 2016. "XGBoost: Scalable Tree Boosting System," *arXiv:1603.02754*.

Dimitrova, D.S., Kaishev, V.K., and Tan, S. 2020. "Computing the Kolmogorov–Smirnov Distribution when the Underlying cdf is Purely Discrete, Mixed or Continuous," *Journal of Statistical Software (95:10)*, pp. 1-42.

Farbmacher, H., Löw, L., and Spindler, M. 2021. "An explainable attention network for fraud detection in claims management," *Journal of Econometrics (228:2),* pp. 244-258.

Hearst, M.A., Dumais, S.T., Osman, E., Platt, J. and Scholkopf, B. 1998. "Support vector machines," *IEEE Intelligent Systems and their Applications (13*), pp. 18-28.

Kaiafas, G., Hammerschmidt, C., State, R., Nguyen, C. D., Ries, T., and Ourdane, M. 2019. "An Experimental Analysis of Fraud Detection Methods in Enterprise Telecommunication Data using Unsupervised Outlier Ensembles," *in Proceedings of the IEEE Symposium on Integrated Network and Service Management,* pp. 37-42.

Kelly, T., Beharry, A., and Fedoruk, M. 2019. "Applying Machine Learning Techniques to Advance Anti-Doping," *European Journal of Sports and Exercise Science (7:2),* pp. 1-9.

Kleppner, D., and Kolenkow, R. 1973. "An Introduction to Mechanics (2nd ed.)," *McGraw-Hill Publications.*

Lever, J., Krzywinski, M. and Altman, N. 2017. "Principal component analysis," *Nature (14),* pp. 641–642.

Mareck, U., Geyer, H., Opfermann, G., Thevis, M. and Schänzer, W. 2008. "Factors influencing the steroid profile in doping control analysis," *Journal of Mass Spectrometer (43),* pp. 877-891.

Massi, M. C., Ieva, F., and Lettieri, E. 2020. "Data mining application to healthcare fraud detection: a two-step unsupervised clustering method for outlier detection with administrative databases," *BMC Medical Informatics and Decision Making (20:160).*

McLaren, R.H. 2016. "WADA investigation of Sochi allegations: The independent person 2nd report," *World Anti-Doping Agency*.

Montagna, S., and Hopker, J. 2018. "A bayesian approach for the use of athlete performance data within anti-doping," *Frontiers in physiology (9),* pp. 884.

Munir, M., Siddiqui, S.A., Dengel, A.R., and Ahmed, S. 2019. "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series," *IEEE Access (7),* pp. 1991-2005.

Nguyen, Q. P., Lim, K. W., Divakaran, D. M., Low, K. H., and Chan, M.C. 2019. "GEE: A Gradient-based Explainable Variational Autoencoder for Network Anomaly Detection," *in Proceedings of the IEEE Conference on Communications and Network Security (CNS),* pp. 91-99.

Óskarsdóttir, M., Ahmed, W., Antonio, K., Baesens, B., Dendievel, R., Donas, T., and Reynkens, T. 2022. "Social Network Analytics for Supervised Fraud Detection in Insurance," *Risk Analysis (42:8),* pp. 1872-1890.

Peng, J., Lee, K, and Ingersoll, G. 2002. "An introduction to Logistic Regression Analysis and Reporting," *Journal of Educational Research (96),* pp. 3-14.

Piper, T., Geyer, H., Haenelt, N., Huelsemann, F., Schaenzer, W., and Thevis, M. 2021. "Current Insights into the Steroidal Module of the Athlete Biological Passport," *International Journal of Sports and Medicine (42),* pp. 1-16.

Psychoula, I., Gutmann, A., Mainali, P., Lee, S., Dunphy, P., and Petitcolas, F. 2021. "Explainable Machine Learning for Fraud Detection," *Computer (54).*

Li, X., Liu, S., Li, Z., Han, X., Shi, C., Hooi, B., Huang, H., and Cheng, X. 2020. "FlowScope: Spotting Money Laundering Based on Graphs," *in Proceedings of the AAAI Conference on Artificial Intelligence, (34:4),*

Pumsirirat, A., and Yan, L. 2018. "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine," *International Journal of advanced computer science and applications (9:1),* pp. 18-25.

Rahman, M.R., Bejder, J., Bonne, T. C., Andersen, A. B., Huertas, J. R., Aikin, R., Nordsborg, N. B., Maass, W. 2022. "Detection of Erythropoietin in Blood to Uncover Doping in Sports using Machine Learning," *in Proceedings of the IEEE International Conference on Digital Health (ICDH),* pp. 193-201.

Renterghem, P.V., Eenoo, P.V., Geyer, H., Schänzer, W., and Delbeke, F.T. 2010. "Reference ranges for urinary concentrations and ratios of endogenous steroids, which can be used as markers for steroid misuse, in a Caucasian population of athletes," *Steroids (75),* pp. 154-163.

Renterghem, P.V., Sottas, P. E., Saugy, M., and Van Eenoo, P. 2013. "Statistical discrimination of steroid profiles in doping control with support vector machines," *Analytica chimica acta (768),* pp. 41-48.

Shmueli, G., and Koppius. O. R. 2011. "Predictive analytics in information systems research," *MIS Quarterly (35:3),* pp. 553–572.

Sidey-Gibbons, J., and Sidey-Gibbons, C. 2019. "Machine learning in medicine: a practical introduction," *BMC Medical Research Methodology (19).*

Sottas, P.-E., Baume, N., Saudan, C., Schweizer, C., Kamber, M., and Saugy, M. 2006. "Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio," *Biostatistics (8:2),* pp. 285-296.

Thevis, M., Geyer, H., Mareck, U., Sigmund, G., Henke, J., Henke, L., and Schänzer, W. 2006. "Detection of manipulation in doping control urine sample collection: a multidisciplinary approach to determine identical urine samples," *Analytical and bioanalytical chemistry (388:7),* pp. 1539–1543.

Thevis, M., Geyer, H., Sigmund, G., and Schänzer, W. 2012. "Sports drug testing: Analytical aspects of selected cases of suspected, purported, and proven urine manipulation," *Journal of pharmaceutical and biomedical analysis (57),* pp. 26–32.

Troncoso, F., and Weber, R. 2020. "A Novel Approach to Detect Associations in Criminal Networks," *Decision Support Systems (128).*

Vellido, A., Martín-Guerrero, J.D., Rossi, F., and Lisboa, P. 2011. "Seeing is believing: The importance of visualisation in real-world machine learning applications," *in the Proceedings of 19th European Symposium on Artificial Neural Networks*.

WADA. 2020. "A summary of WADA investigations into the International Weightlifting Federation and the sport of weightlifting. Intelligence and Investigation Department," *World Anti-Doping Agency. URL: https://www.wada-ama.org/sites/default/files/resources/files/iwf_global_report.pdf* (visited on May 03, 2022).

WADA. 2021. "Athlete Passport Management Unit Requirements and Procedures," *WADA Technical Document – TD2021APMU. URL: https://www.wada-ama.org/sites/default/files/resources/files/td2021apmu_final_eng_0.pdf* (visited on May 03, 2022).

WADA. 2021. "Decision limits for the confirmatory quantification of exogenous threshold substances by chromatography-based analytical methods," *WADA Technical Document – TD2021DL. URL: https://www.wada-ama.org/sites/default/files/resources/files/td2021dl_final_eng_0.pdf* (visited on May 03, 2022).

WADA. 2021. "International Standard for the Protection of Privacy and Personal Information (ISPPPI)," *URL: https://www.wada-ama.org/sites/default/files/resources/files/2021_ispppi_en.pdf* (visited on May 03, 2022).

WADA. 2021. "Measurement and Reporting of Endogenous Anabolic Androgenic Steroid (EAAS) Markers of the Urinary Steroid Profile," *WADA Technical Document – TD2021EAAS. URL: https://www.wada-ama.org/sites/default/files/resources/files/td2021eaas_final_eng_v_2.0.pdf* (visited on May 03, 2022).

WADA. 2021. "World Anti-Doping Code 2021," *World Anti-Doping Agency. URL: https://www.wada-ama.org/sites/default/files/resources/files/2021_wada_code.pdf* (visited on May 03, 2022).

Wilkes, E. H., Rumsby, G., and Woodward, G. M. 2018. "Using machine learning to aid the interpretation of urine steroid profiles," *Clinical chemistry (64:11),* pp. 1586-1595.

Zheng, W., Yan, L., Gou, C., and Wang, F.-Y. 2021. "Federated meta-learning for fraudulent credit card detection," *in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (642),* pp. 4654–4660.