

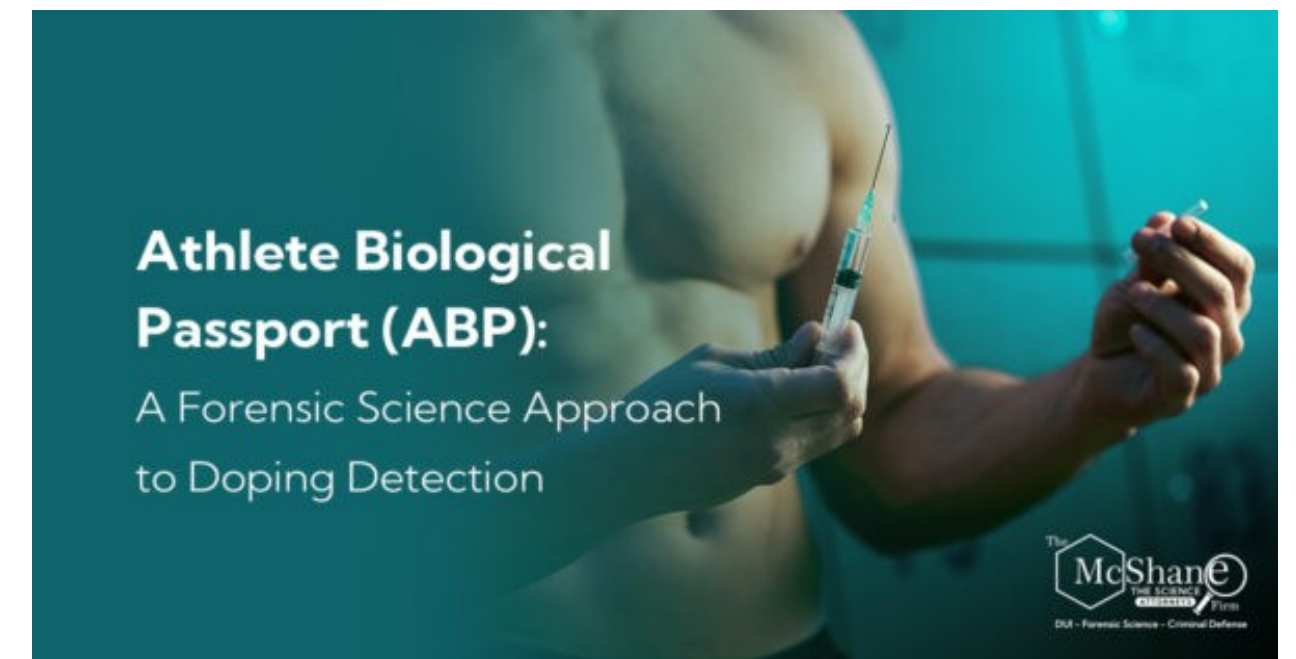
Identification of doping suspicions through artificial intelligence-powered analysis on athlete's performance passport in female weightlifting

SDA 정하늘

DOPING

스포츠에서의 도핑

- 스포츠 경기 내 공정 경쟁의 근본 원칙을 훼손 → 엄격히 금지
- 도핑 방지 규정에 대한 인식 부족을 시정하고
도핑 통제를 강화함을 목적으로 조치 시행
→ ABP(Athlete Biological Passport)
→ 도핑 방지 규정 위반(ADRVs)을 식별하는 표준 검사로 부상



ABP's Problem

prescribed medications



individual variations



detection window



delayed yet sustained performance

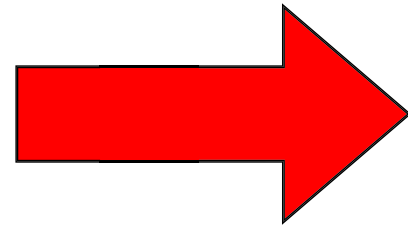


Athlete's Performance Passport(APP)

: 인구통계학적 프로파일과 퍼포먼스 데이터를 포함하는 선수 퍼포먼스가 있는 일종의 기록지

→ 비정상적인 기량 향상 및/또는 지속적인 고성능 기록을 식별할 기회를 제공

통계 분석과 AI 알고리즘을 결합
도핑 방지 관행에 APP를 활용



깨끗한 집단과 도핑 집단의 차이 발견
퍼포먼스의 휘발성을 결정하는 통계 모델 도입
비정상적인 퍼포먼스 향상을 식별할 수 있는 모델 도입

APP에 기반하여 의심 선수를 표적화하는 정확한 모델의 개발

→ 도핑 의심 개인을 선정하고 검사하기 위한 기준 확립의 보조적 증거를 제공

역도는 왜 도핑에 더욱 관심을 깊게 가지는가

1. 올림픽 속 논란

2008 베이징 올림픽 & 2012 런던 올림픽
515명 참가자 중 30명의 역도 선수가 메달 소급 박탈

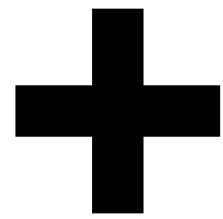
도핑의 오랜 역사를 가진 국가들에서
도핑 방지 규정 위반(ADRVs)에 대한 충분한 조치가
시행되었는지에 대한 의문 제기



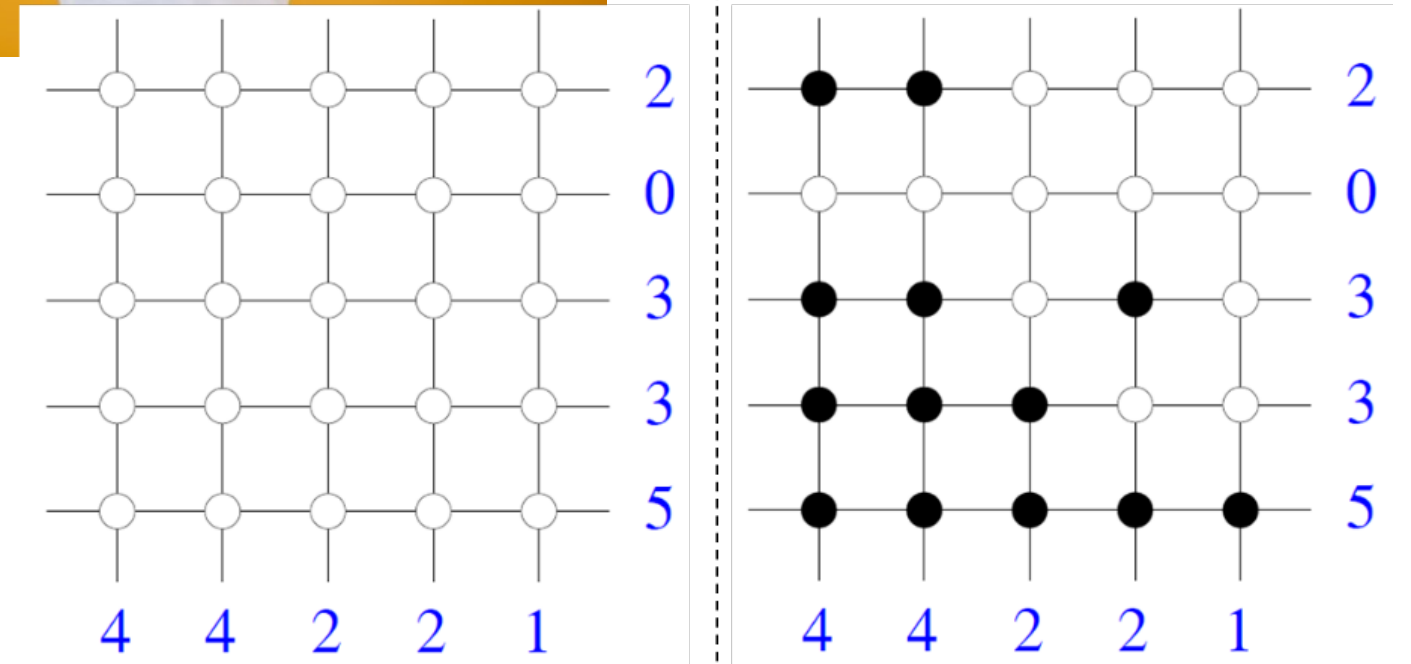
역도는 왜 도핑에 더욱 관심을 깊게 가지는가

2. 역도라는 스포츠 특성

역도: 속도와 힘 모두에 프리미엄을 두는 스포츠
선수의 체중에 의해 크게 영향



역도: 이산적인 수치로 정확하게 정량화되는 스포츠



Data acquisition and processing

국제역도연맹 사이트에 존재하는 공식 기록 이용
연세대학교 생명윤리심의위원회(IRB)에게 승인 면제

→ 1998년부터 2020년까지 IWF가 주관한 모든 대회의
여성 역도 선수 인구통계 데이터와 퍼포먼스 결과 포함

인구통계 데이터: 성별, 연령, 체중(BW), 도핑 이력 등

퍼포먼스 결과: 인상(Snatch)과 용상(Clean and Jerk)의 최고 성적 합계인 총 중량
(각각은 세 번의 시도로 구성)

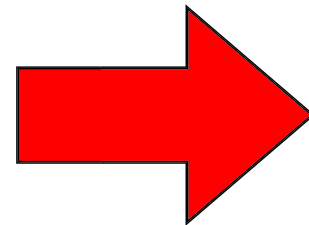


INTERNATIONAL
WEIGHTLIFTING
FEDERATION



Data acquisition and processing

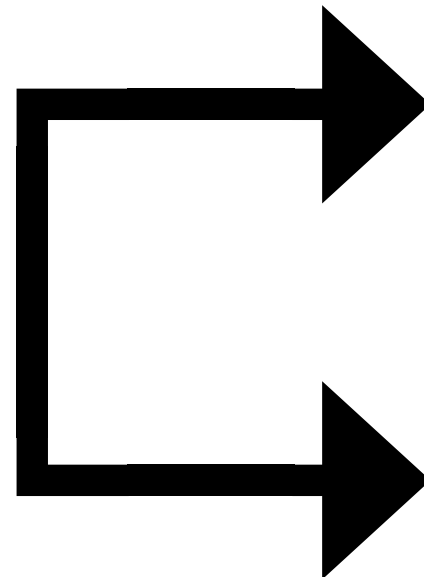
총 19,591 기록 획득



17,058개 산출

퍼포먼스 기록, 연령, 체중 정보 無
2,533개의 엔티티 제거

17,058개 중



15,404개 - 징계 이력이 없는 선수(비징계 그룹)

1,654개 - 징계 이력이 있는 선수(징계 그룹)

categorized

세 가지 파라미터: 연령대, 체중급, 퍼포먼스 수준

연령대	15세 미만	15 ~ 19	20 ~ 24	25 ~ 29	30 ~ 34	35 ~ 39	40세 이상	
체중급	49	55	59	64	76	87	87+	
퍼포먼스	상위 1%	상위 1%-5%	상위 5%-10%	상위 10%-25%	상위 25%-50%	상위 50%-75%	상위 75%-90%	상위 90%-100%

Methods: XGBoost + MLP, Ensemble

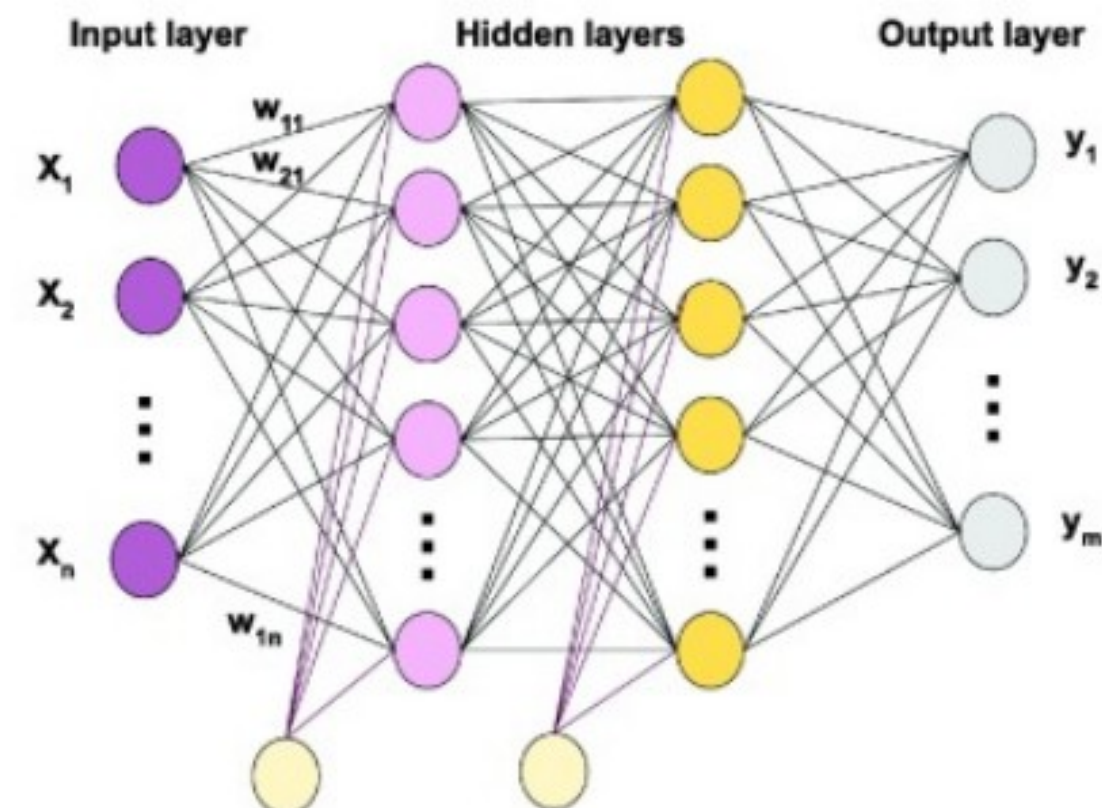
Five feature: 연령, 체중, 인상기록, 용상기록, 개인의 해당 체급

추가 정제 -

연령: 15세 미만 제외

체중: 0에서 6까지의 서수형 변수로 인코딩, 65 미만 & 125 이상 제외

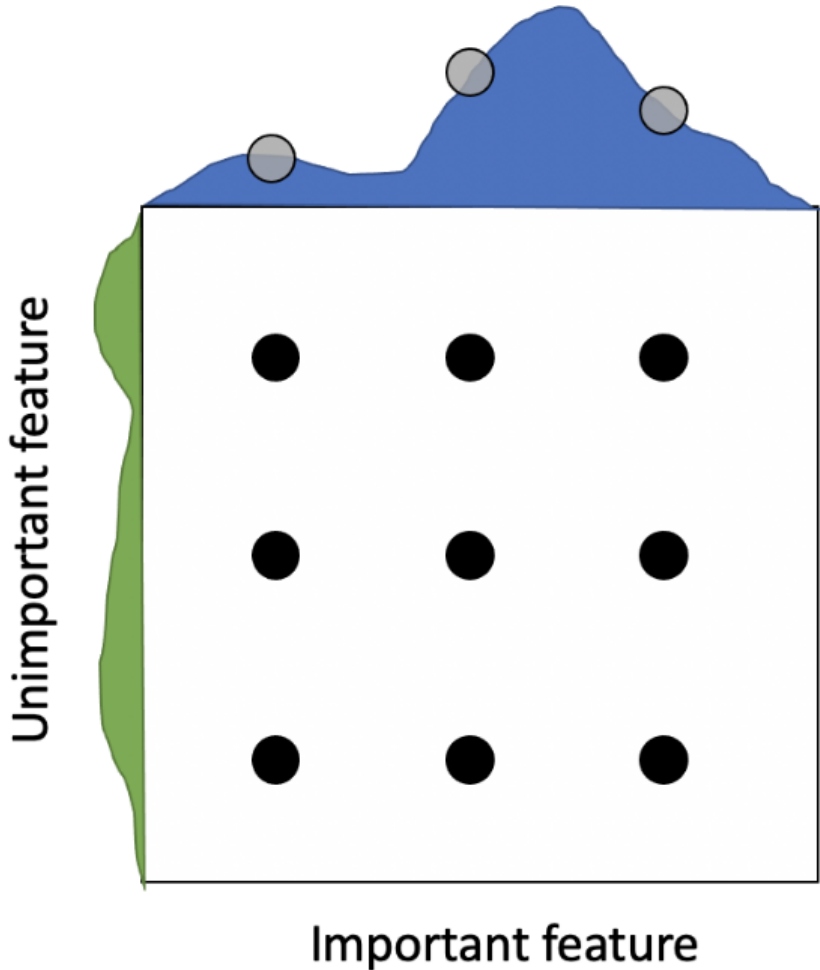
⇒ 8,948개의 엔티티로 구성된 훈련 세트 생성



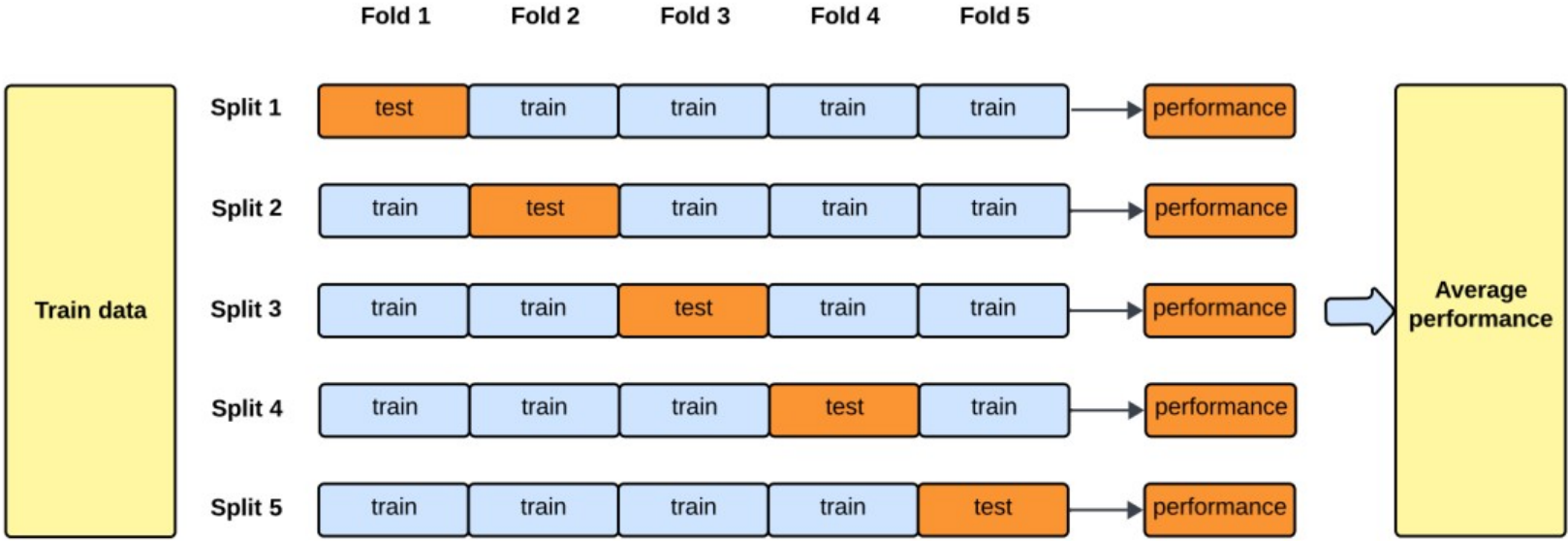
advancement

grid search

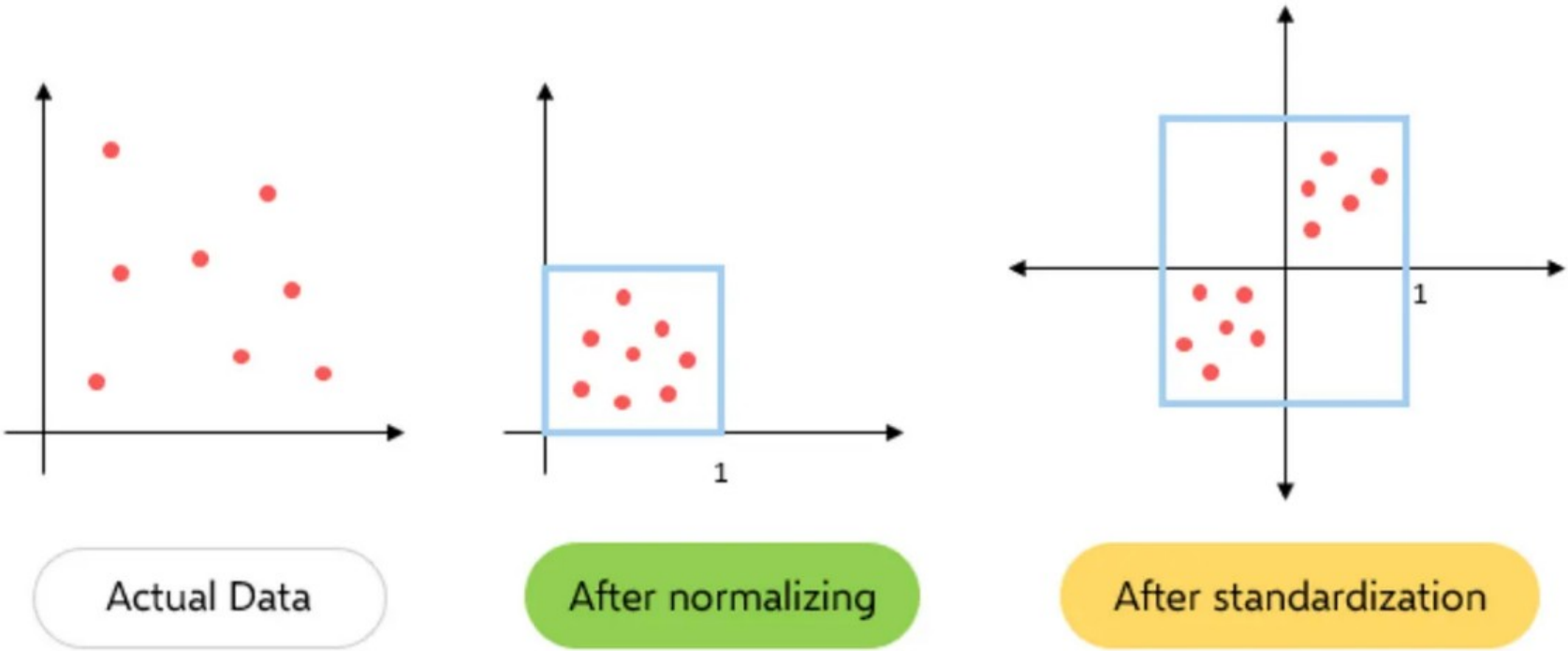
Grid Search
with 3x3 grid



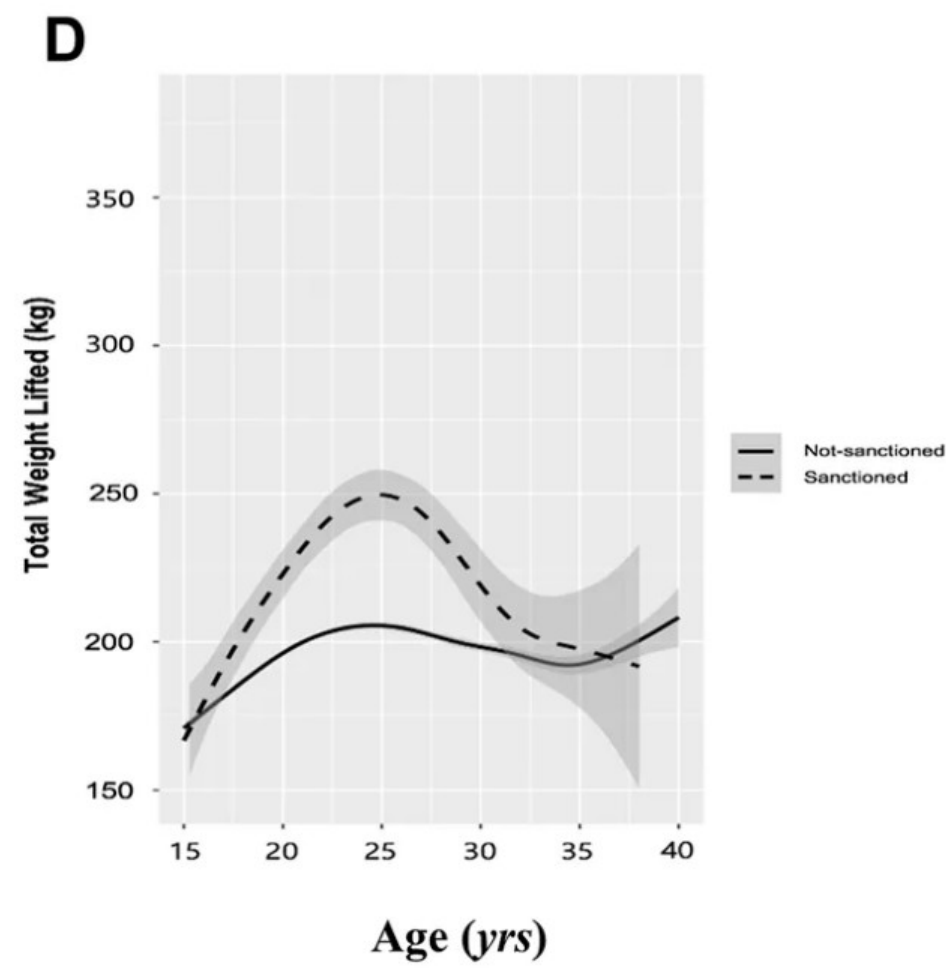
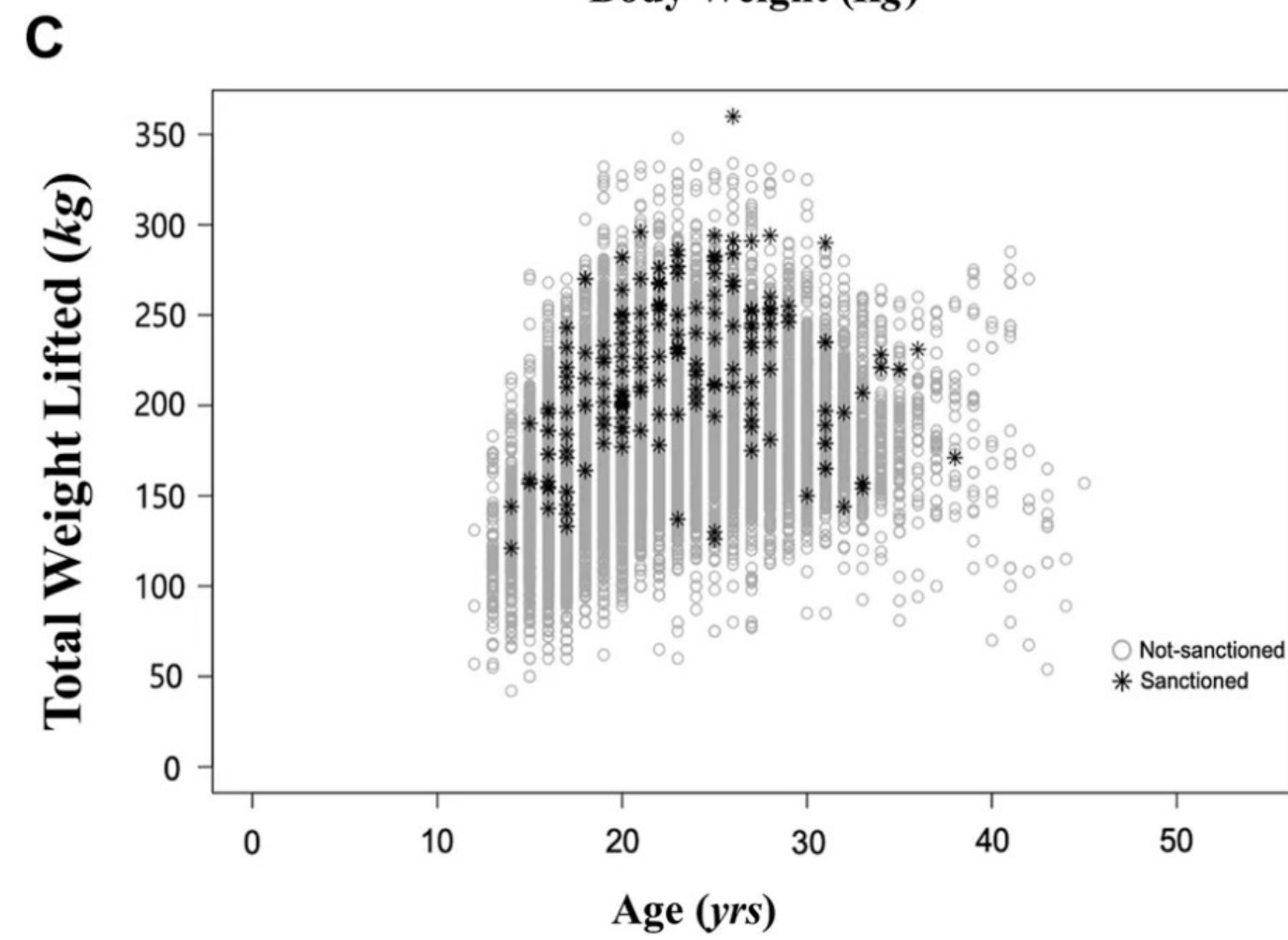
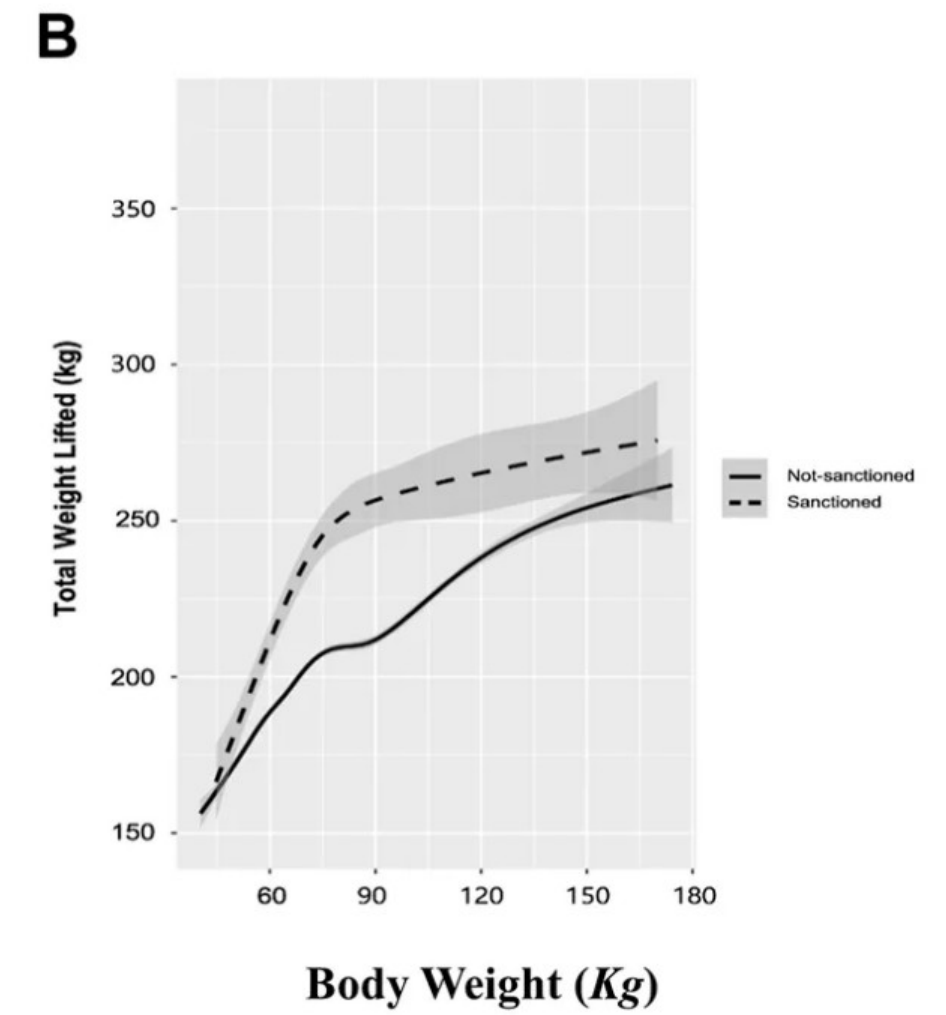
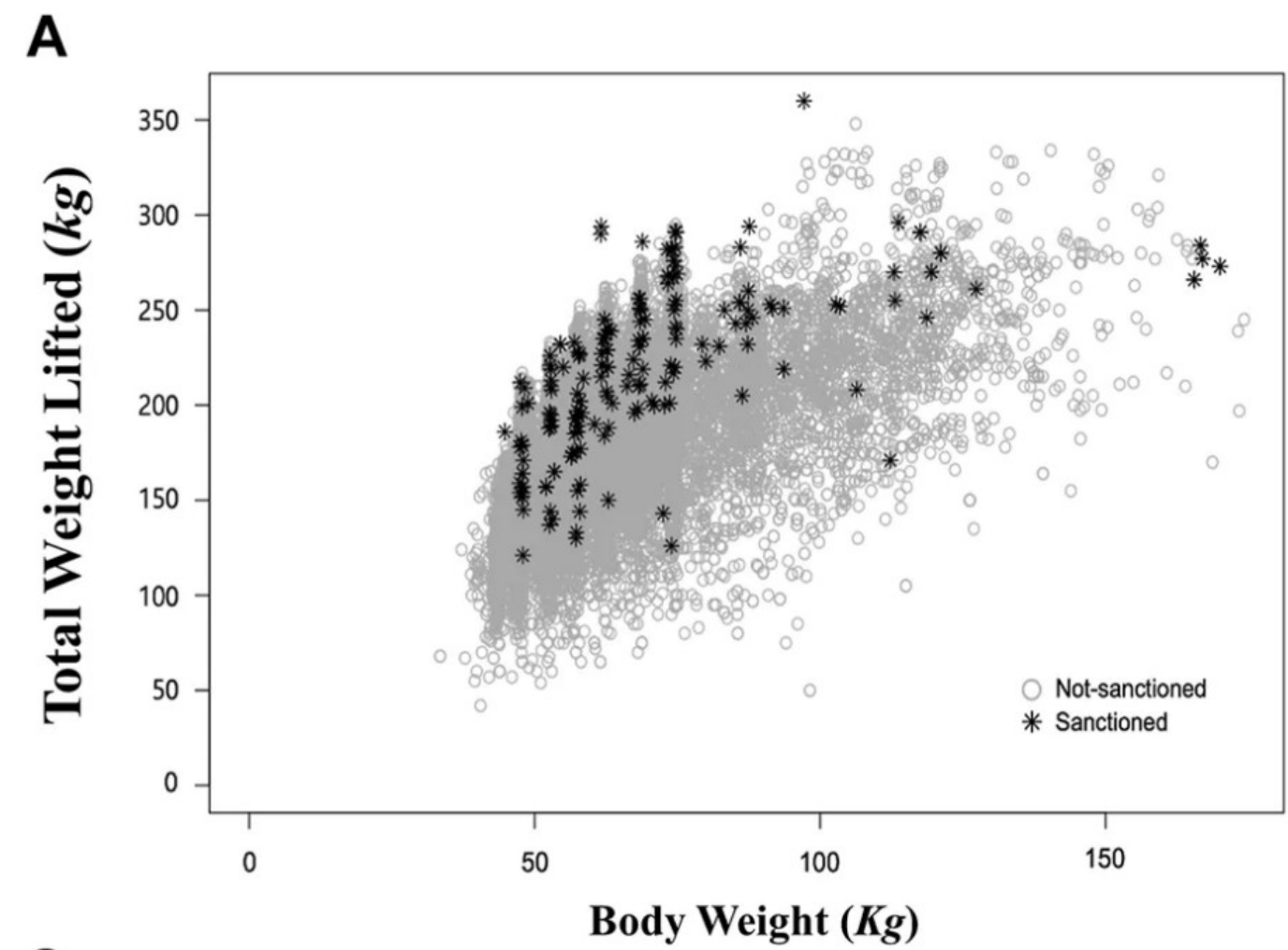
5-fold cross-validation



data standardization



Result



Result

TABLE 1 Distribution of sanctioned cases across age groups and performance levels in female weightlifting analysis of 17,058 cases.

Age group	Top 1% (<i>n</i> = 155)	Top 1%–5% (<i>n</i> = 681)	Top 5%–10% (<i>n</i> = 823)	Top 10%–25% (<i>n</i> = 2,490)	Top 25%–50% (<i>n</i> = 4,174)	Top 50%–75% (<i>n</i> = 4,219)	Top 75%–90% (<i>n</i> = 2,613)	Top 90%–100% (<i>n</i> = 1903)	ADRVs
<15 (<i>n</i> = 81)	0	1	0	4	4	4	1	0	14 (17.3%)
15–19 (<i>n</i> = 8,636)	24	59	58	96	137	98	26	7	505 (5.9%)
20–24 (<i>n</i> = 4,834)	11	72	63	111	137	114	52	11	571 (11.8%)
25–29 (<i>n</i> = 2,485)	4	42	32	66	105	79	41	19	388 (15.6%)
30–34 (<i>n</i> = 818)	3	12	12	24	26	36	26	9	148 (18.1%)
35–39 (<i>n</i> = 161)	0	0	0	4	5	12	4	1	26 (16.2%)
≥40 (<i>n</i> = 34)	0	0	0	0	1	1	0	0	2 (5.9%)
ADRVs	42 (27.1%)	186 (27.3%)	165 (20.1%)	305 (12.3%)	415 (9.9%)	344 (8.2%)	150 (5.7%)	47 (2.5%)	1,654

*The percentages in the parentheses denote the number of ADRVs, per analyzed cases in the matched age group (rows) or in the matched performance level (columns).

Result

TABLE 2 Distribution of sanctioned cases across body weight classes and performance levels in female weightlifting analysis of 17,058 cases.

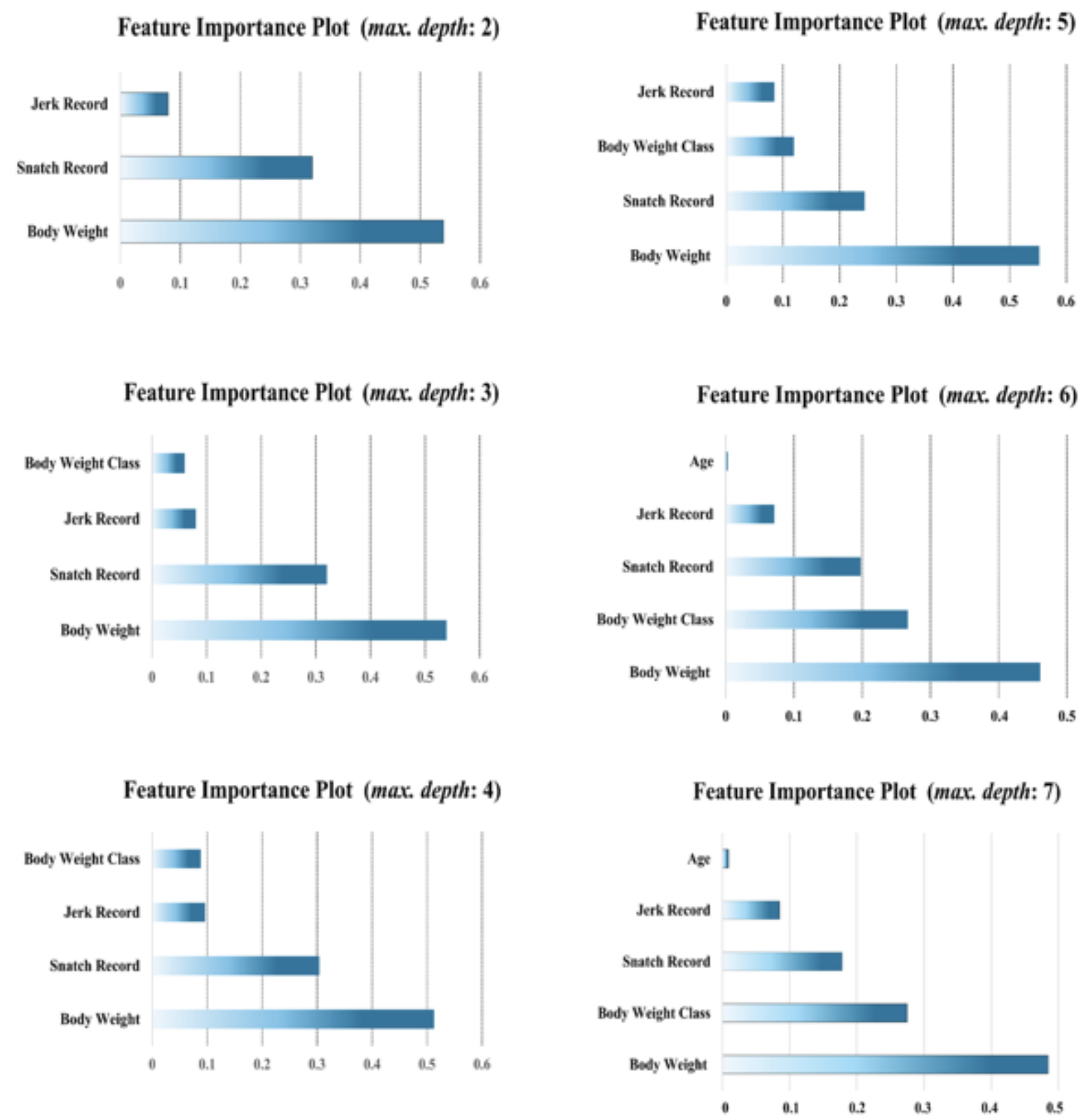
BW class	Top 1% (<i>n</i> = 155)	Top 1%–5% (<i>n</i> = 681)	Top 5%–10% (<i>n</i> = 823)	Top 10%–25% (<i>n</i> = 2,490)	Top 25%–50% (<i>n</i> = 4,174)	Top 50%–75% (<i>n</i> = 4,219)	Top 75%–90% (<i>n</i> = 2,613)	Top 90%–100% (<i>n</i> = 1903)	ADRVs
49 (<i>n</i> = 2,642)	9	21	21	65	59	33	10	1	219 (8.3%)
55 (<i>n</i> = 2,539)	5	23	23	62	74	48	14	2	251 (9.9%)
59 (<i>n</i> = 2,628)	4	24	18	46	55	58	17	4	226 (8.6%)
64 (<i>n</i> = 2,678)	12	31	22	70	53	39	7	3	237 (8.9%)
76 (<i>n</i> = 4,007)	23	60	44	96	103	66	25	10	427 (10.7%)
87 (<i>n</i> = 813)	2	10	4	14	15	6	0	0	51 (6.3%)
87+ (<i>n</i> = 1751)	4	16	32	69	79	36	6	1	243 (13.9%)
ADRVs	59 (38.1%)	185 (27.5%)	164 (20.5%)	422 (16.7%)	438 (10.6%)	286 (6.8%)	79 (3.0%)	21 (1.1%)	1,654

Result

TABLE 3 Performance metrics of AI-Powered models for doping suspicion prediction in female weightlifting.

Model		Train (5-fold CV)			Test		
		Accuracy	AUC-ROC	F1 score	Accuracy	AUC-ROC	F1 score
Logistic Regression		0.710	0.695	0.300	0.852	0.761	0.581
XGBoost	Depth: 2	0.844	0.636	0.254	0.784	0.685	0.457
	Depth: 3	0.710	0.6623	0.285	0.796	0.730	0.471
	Depth: 4	0.773	0.677	0.305	0.796	0.746	0.471
	Depth: 5	0.815	0.689	0.308	0.830	0.777	0.546
	Depth: 6	0.798	0.694	0.315	0.875	0.790	0.621
	Depth: 7	0.818	0.695	0.321	0.864	0.780	0.600
MLP	Hidden: 2	0.764	0.701	0.305	0.830	0.748	0.546
	Hidden: 3	0.772	0.705	0.311	0.875	0.760	0.621
	Hidden: 4	0.716	0.709	0.315	0.886	0.766	0.615
Ensemble	Depth: 5 Hidden: 4	0.693	0.697	0.313	0.875	0.783	0.645

CV, cross validation; MLP, multilayer perceptron; AUC-ROC, area under the curve of receiver operating characteristics.



Result

TABLE 4 Ensemble model performance in predicting doping suspicions among female weightlifters–2008, 2012, 2016 olympics.

Best prediction by ensemble		References	
		Not -sanctioned	Sanctioned
2008 Beijing	Not-Sanctioned	67	5
	Sanctioned	6	10
2012 London	Not-Sanctioned	61	7
	Sanctioned	13	9
2016 Rio de Janeiro	Not-Sanctioned	79	6
	Sanctioned	4	2

Discussion

25.6세의 법칙: 생화학적으로 여성의 근육량과 폭발력 - 25.6세에 정점을 찍고 자연스럽게 감소

1. 20대 초반에 최고 성적을 내야 한다는 압박감
2. 30대에 들어서며 나타나는 노화를 약물(화학적 수단)로 강제 억제
'평균보다 훨씬 부풀려진(Inflated) 물리적 출력'을 가진다



Discussion

1

데이터
순도

2

연속적
스포츠
에서의
적용

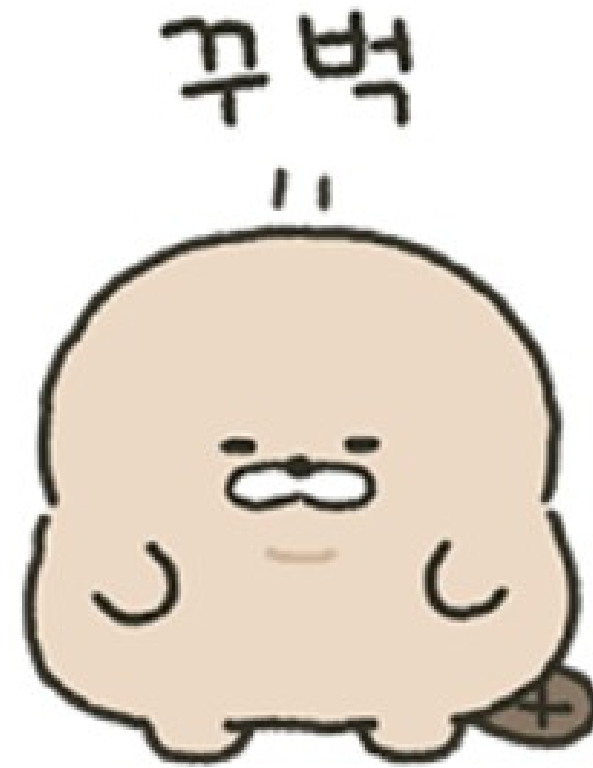
3

약물과
선수 개인
노력 구별
미비

4

예측률
하락

발표 마치겠습니다.



들어주셔서 감사합니다.