FROM PIXEL TO PODIUMS:

Opportunities and Challenges for Deep Learning in Sports

Leonid Kholkine

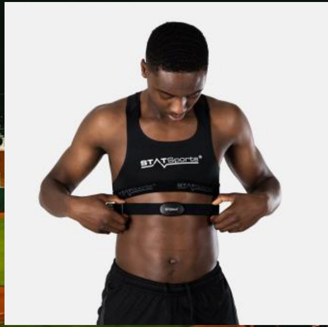
IDLab, imec research group at Ghent University and University of Antwerp

Sports is a Jungle of Data

Sports is a Jungle of Data

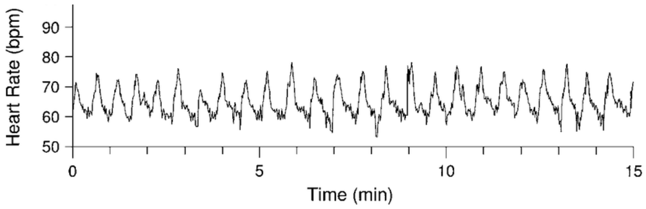
Sports is a Jungle of Data

Sports is a Jungle of Data

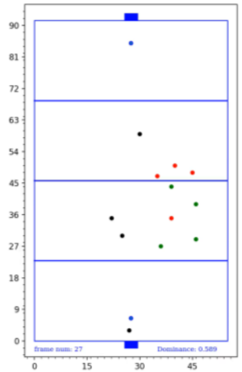


What can it be used for?

Typical data used in sports



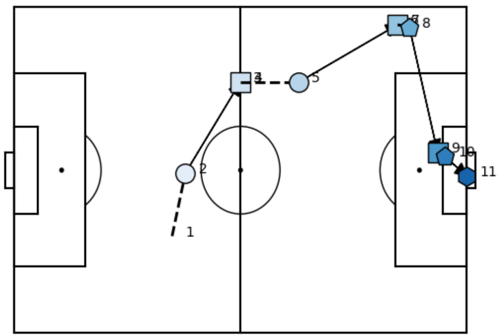
Sensor time series



Spatial-Temporal

	Driver	Number	Team	Grid	Pits	Fastest Lap	Race Time
1	Max Verstappen	33	Red Bull	3	1	1:18.999	1:38:39.086
2	Lewis Hamilton	44	Mercedes	2	1	1:19.820	+16.555
3	Sergio Perez	11	Red Bull	4	1	1:19.468	+17.752
4	Pierre Gasly	10	AlphaTauri	5	1	1:20.510	+1:03.845
5	Charles Leclerc	16	Ferrari	8	1	1:20.665	+1:21.037
6	Carlos Sainz Jnr	55	Ferrari	6	1	1:20.081	+1 lap
7	Sebastian Vettel	5	Aston Martin	9	1	1:20.460	+1 lap
8	Kimi Raikkonen	7	Alfa Romeo	10	1	1:20.713	+1 lap
9	Fernando Alonso	14	Alpine	12	1	1:20.711	+1 lap
10	Lando Norris	4	McLaren	18	1	1:20.617	+1 lap
11	Antonio Giovinazzi	99	Alfa Romeo	11	1	1:21.523	+1 lap
12	Daniel Ricciardo	3	McLaren	7	2	1:21.069	+1 lap
13	Esteban Ocon	31	Alpine	19	1	1:21.348	+1 lap
14	Lance Stroll	18	Aston Martin	20	2	1:20.930	+2 laps
15	Valtteri Bottas	77	Mercedes	1	4	1:17.774	+2 laps
16	George Russell	63	Williams	16	1	1:22.016	+2 laps
17	Nicholas Latifi	6	Williams	13	2	1:21.546	+2 laps
18	Nikita Mazepin	9	Haas	15	2	1:21.402	+3 laps
NC	Yuki Tsunoda	22	AlphaTauri	17	0	-	DNF (0)
NC	Mick Schumacher	47	Haas	14	0	-	DNF (0)

Open Data

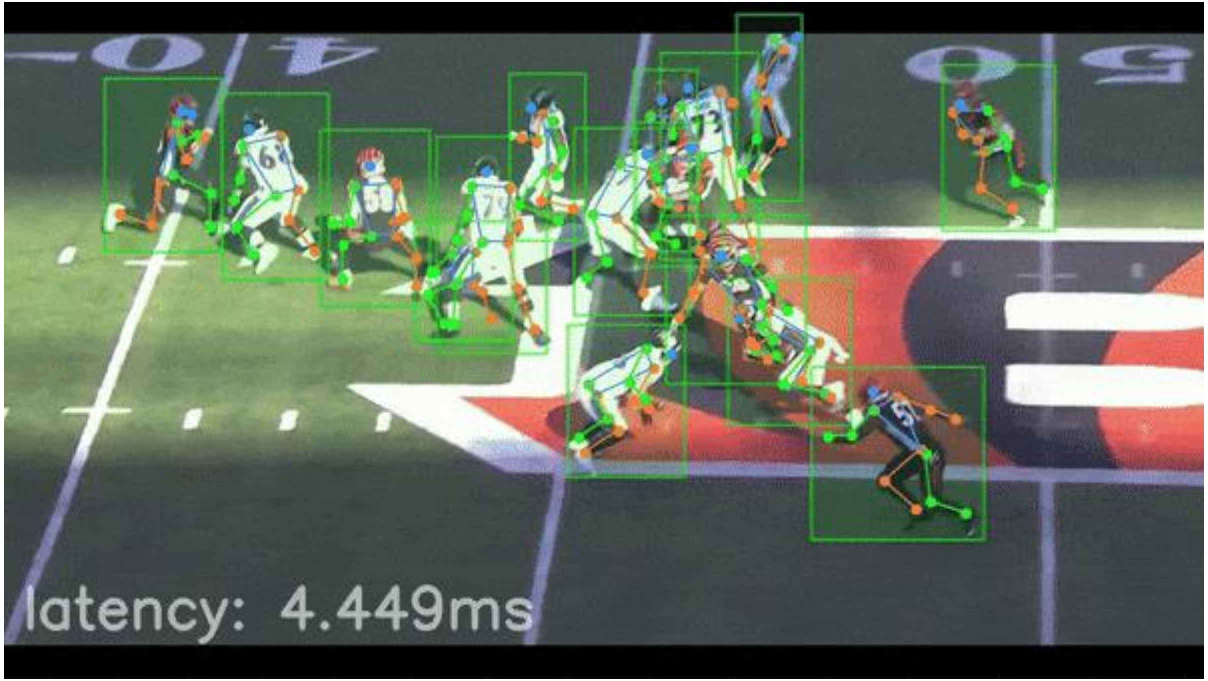


Events

Video

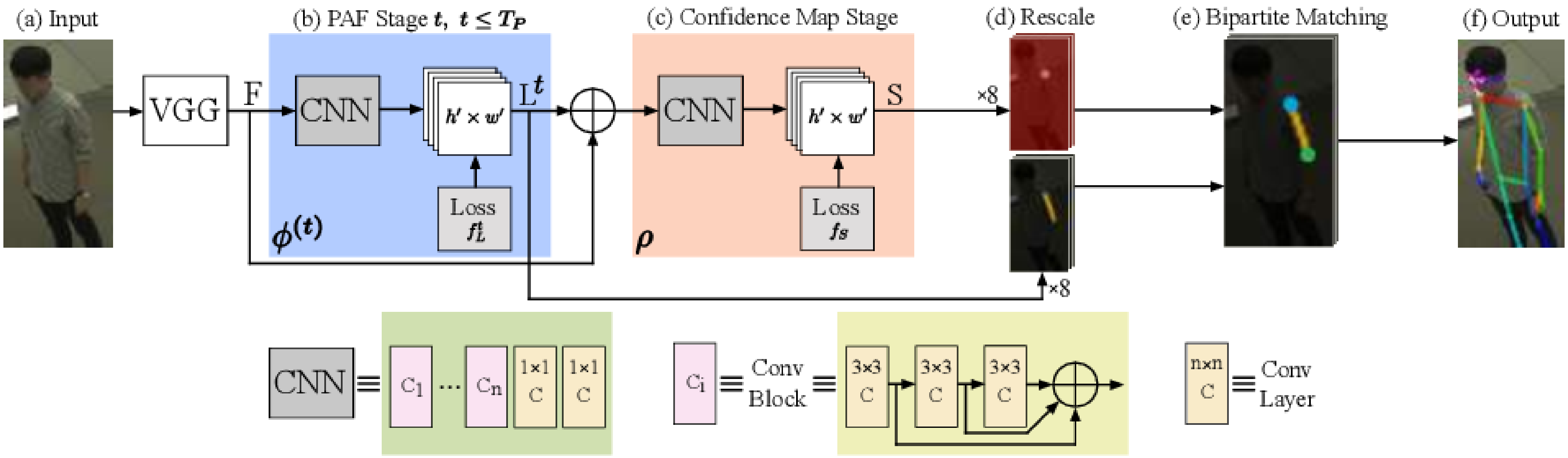
Video Data

Pose Estimation



RTMO: Towards High-Performance One-Stage Real-Time Multi-Person Pose Estimation, Peng Lu and Tao Jiang and Yining Li and Xiangtai Li and Kai Chen and Wenming Yang

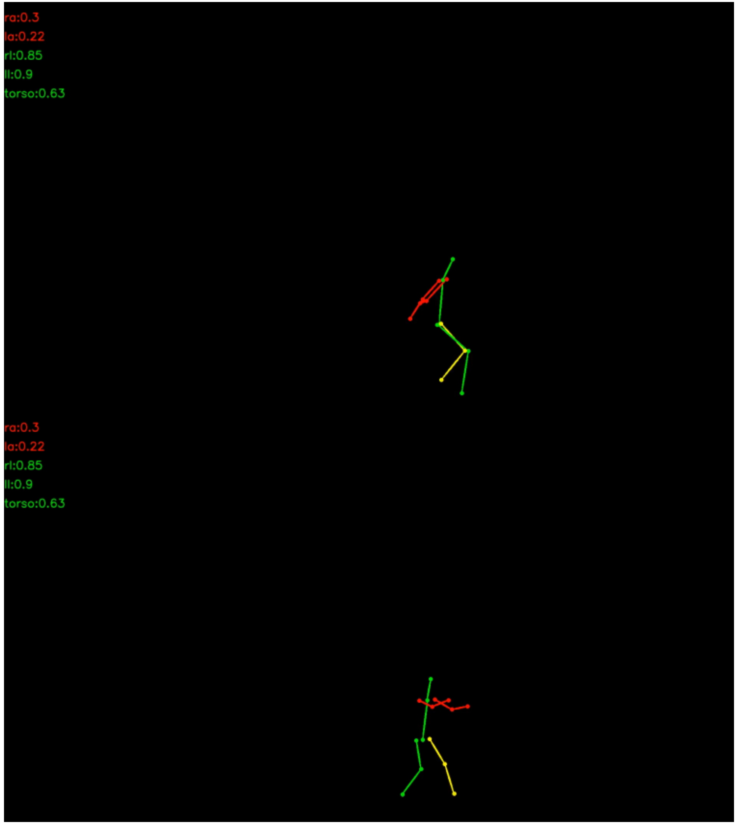
Pose Estimation



OpenPose : Whole-Body Pose Estimation
Gines Hidalgo, Yaser Sheikh, Kris Kitani, Aayush Bansal, Ramon Sanabria, Donglai Xiang, Xiu Li, Haroon Idrees

Automated Feedback from Pose Estimation in Volleyball

[Unpublished] Jakub Kaczmarek, Leonid Kholkine, Dietrich Heiser



Tracking Systems

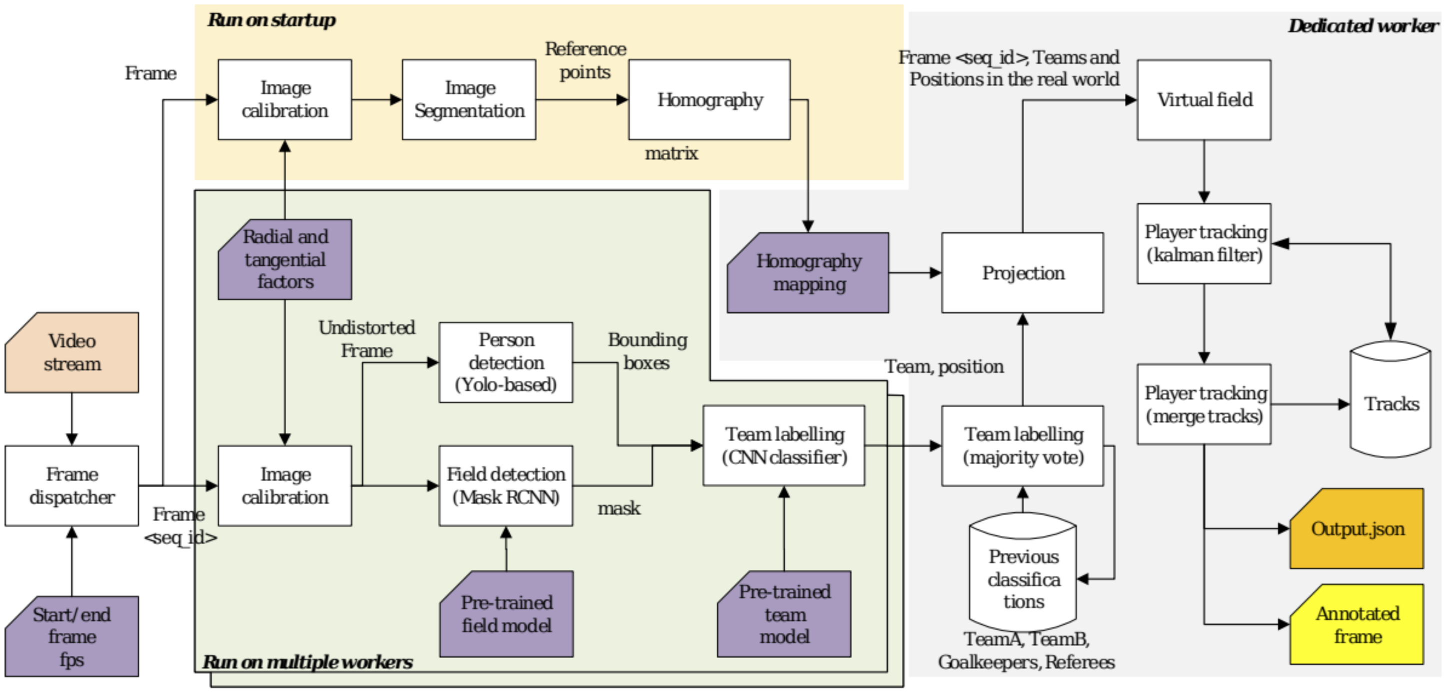
Low-cost player tracking in field hockey

Henrique Duarte Moura, Leonid Kholkine, Laurens Van Damme, Kevin Mets, Christiaan Leysen, Tom De Schepper, Peter Hellinckx, Steven Latré

- Detect players in the pitch
- Low cost setup
- Output:
  - Json with players detected (ids, teams, position, speed)
  - Virtual field as image

Low-cost player tracking in field hockey

Henrique Duarte Moura, Leonid Kholkine, Laurens Van Damme, Kevin Mets, Christiaan Leysen, Tom De Schepper, Peter Hellinckx, Steven Latré



imec IDLab UNIVERSITEIT GENT Universiteit Antwerpen

Low-cost player tracking in field hockey

Henrique Duarte Moura, Leonid Kholkine, Laurens Van Damme, Kevin Mets, Christiaan Leysen, Tom De Schepper, Peter Hellinckx, Steven Latré

RESULTS

- Player Detection: 79.3% F1 for YoLo v5s
- Field Detection: 72.8% accuracy
- Team Labeling: 98% accuracy
- System accuracy: 93.8% accuracy

- Where it fails:
  - Occlusions
  - Duplications
  - Unrecognizable player

Tracking Data

Why track?

- Automated Event Labeling
- Automated Refereeing
- Automated Tactical Analysis

Data-Driven Ghosting using Deep Imitation Learning

HoangM. Le, Peter Carr, Yisong Yue, and Patrick Lucey

- Answering the problem of "What If?"
  - What would have happened in this situation for the team?
  - What about for the league?
- How? Two stage LSTM Model
  - Stage 1: Learn average behavior of each player in the defending team
  - Stage 2: Use pre-trained model to scale-up into a multi-agent learning

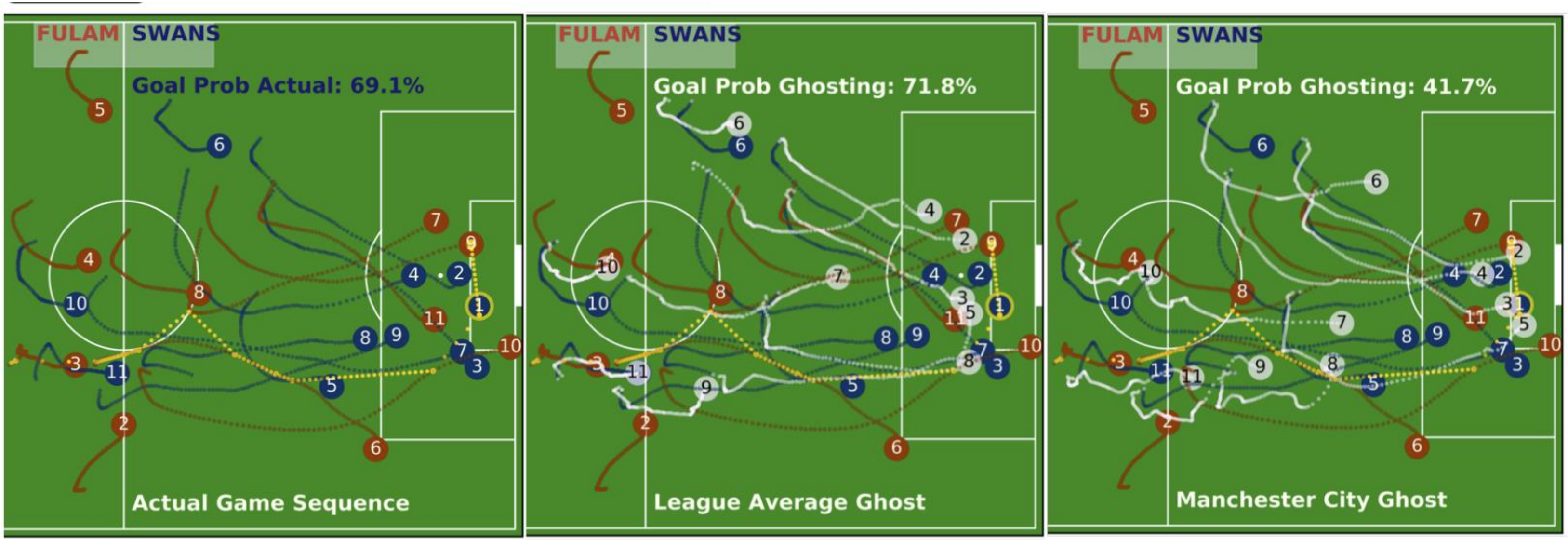
https://www.youtube.com/watch?v=130r0XQIKq0

http://www.yisongyue.com/publications/ssac2017_ghosting.pdf

https://www.youtube.com/watch?v=WI-WL2cj0CA

Data-Driven Ghosting using Deep Imitation Learning

HoangM. Le, Peter Carr, Yisong Yue, and Patrick Lucey

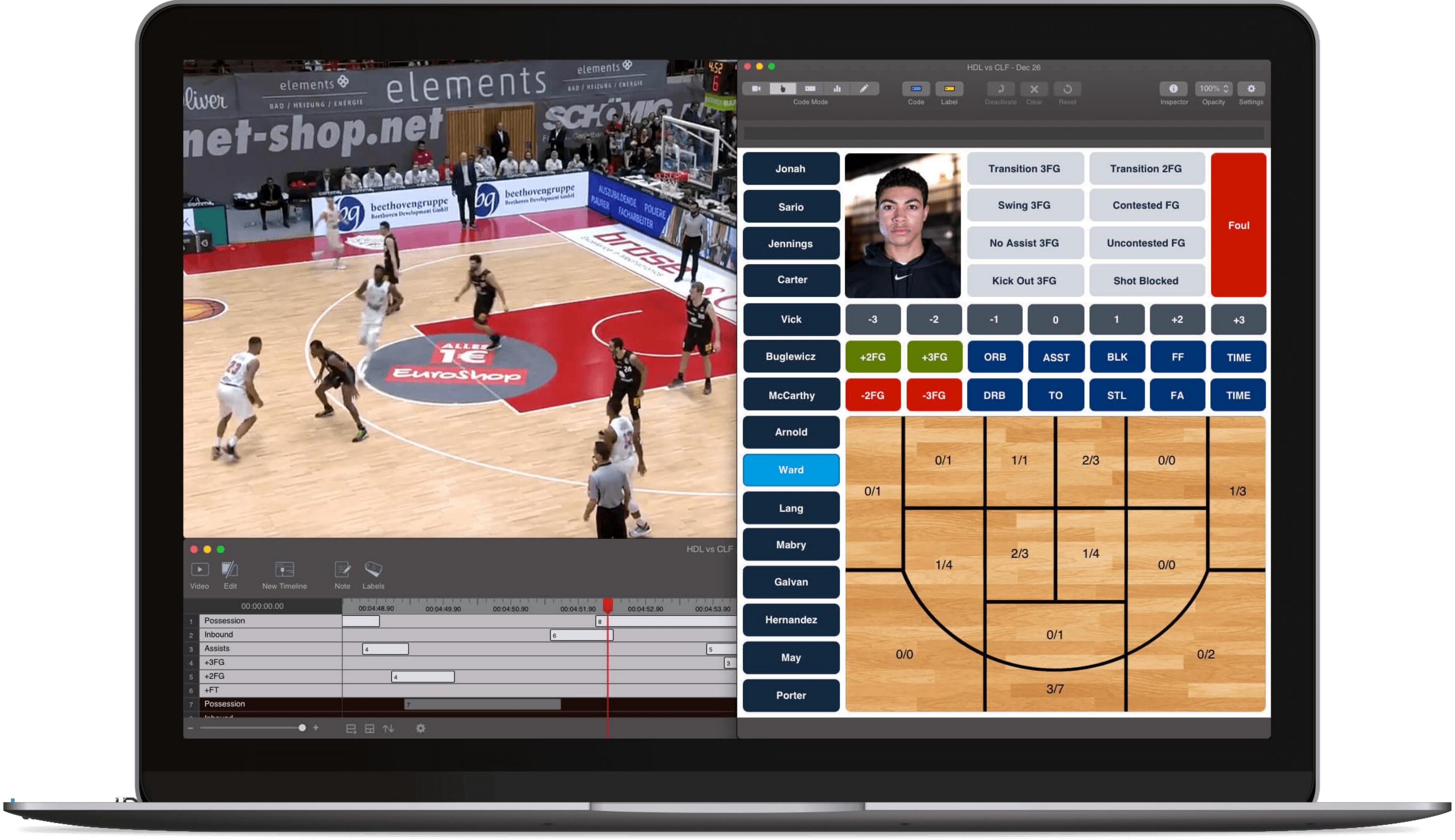


https://www.youtube.com/watch?v=130r0XQIKq0

http://www.yisongyue.com/publications/ssac2017_ghosting.pdf

https://www.youtube.com/watch?v=WI-WL2cj0CA

Event Data


HDL vs CLF - Dec 26
Code Mode
Code
Label
Deactivate
Clear
Reset
Inspector
100%
Opacity
Settings
Jonah
Sario
Jennings
Carter
Vick
Buglewicz
McCarthy
Arnold
Ward
Lang
Mabry
Galvan
Hernandez
May
Porter
Transition 3FG
Transition 2FG
Swing 3FG
Contested FG
No Assist 3FG
Uncontested FG
Kick Out 3FG
Shot Blocked
Foul
-3
-2
-1
0
1
+2
+3
+2FG
+3FG
ORB
ASST
BLK
FF
TIME
-2FG
-3FG
DRB
TO
STL
FA
TIME
0/1
1/1
2/3
0/0
0/1
1/3
1/4
2/3
1/4
0/0
0/1
0/0
0/2
3/7
HDL vs CLF
Video
Edit
New Timeline
Note
Labels
00:00:00.00
00:04:48.90
00:04:49.90
00:04:50.90
00:04:51.90
00:04:52.90
00:04:53.90
Possession
Inbound
Assists
+3FG
+2FG
+FT
Possession
elements
net-shop.net
beethovengruppe
EuroShop

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

How do you evaluate players, not just based on goals?

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

How do you evaluate players, not just based on goals?

All actions in football have the goal to either score a goal or defend it.

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

How do you evaluate players, not just based on goals?

All actions in football have the goal to either score a goal or defend it.

$$\Delta P_{score}(a_i, x) = P_{score}(s_i, x) - P_{score}(s_{i-1}, x)$$

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

How do you evaluate players, not just based on goals?

All actions in football have the goal to either score a goal or defend it.

$$\Delta P_{score}(a_i, h) = P_{score}(s_i, x) - P_{score}(s_{i-1}, x)$$

$$\Delta P_{concedes}(a_i, h) = P_{concedes}(s_i, x) - P_{concedes}(s_{i-1}, x)$$

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

How do you evaluate players, not just based on goals?

All actions in football have the goal to either score a goal or defend it.

$$\Delta P_{score}(a_i, h) = P_{score}(s_i, x) - P_{score}(s_{i-1}, x)$$

$$\Delta P_{concedes}(a_i, h) = P_{concedes}(s_i, x) - P_{concedes}(s_{i-1}, x)$$

$$V(a_i, x) = \Delta P_{score}(a_i, h) + (-\Delta P_{concedes}(a_i, h))$$

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

How do you evaluate players, not just based on goals?

All actions in football have the goal to either score a goal or defend it.

$$\Delta P_{score}(a_i, h) = P_{score}(s_i, x) - P_{score}(s_{i-1}, x)$$

$$\Delta P_{concedes}(a_i, h) = P_{concedes}(s_i, x) - P_{concedes}(s_{i-1}, x)$$

$$V(a_i, x) = \Delta P_{score}(a_i, h) + (-\Delta P_{concedes}(a_i, h))$$

↓

Probabilistic Classifier (XGBoost, Neural Network, …)

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

How do you evaluate players, not just based on goals?

- StartTime: the action's start time,
- EndTime: the action's end time,
- StartLoc: the (x,y) location where the action started,
- EndLoc: the (x,y) location where the action ended,
- Player: the player who performed the action,
- Team: the player's team,
- ActionType: the type of the action (e.g., pass, shot, dribble),
- BodyPart: the player's body part used for the action,
- Result: the result of the action (e.g., success or fail).

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis

	TIME	PLAYER	ACTION	P_{scores}	VALUE
1	92m4s	S. Busquets	pass	0.03	0.00
2	92m6s	L. Messi	pass	0.02	- 0.01
3	92m8s	S. Busquets	pass	0.03	+ 0.01
4	92m11s	L. Messi	take on	0.08	+ 0.05
5	92m12s	L. Messi	pass	0.17	+ 0.09
6	92m14s	A. Vidal	shot	1.00	+ 0.83

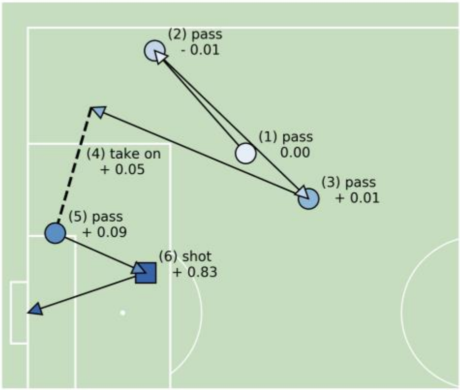


Figure 1: The attack leading up to Barcelona's final goal in their 3-0 win against Real Madrid on December 23, 2017.

R_{vaep}	Player	Rating	R_g	R_a	R_{g+a}	Market Value
1	P. Coutinho	0.899	10	2	4	€ 140m
2	M. Salah	0.817	1	23	2	€ 150m
3	K. De Bruyne	0.641	72	4	15	€ 150m
4	E. Hazard	0.636	21	122	34	€ 150m
5	R. Mahrez	0.635	34	11	16	€ 60m
6	A. Martial	0.607	13	13	9	€ 60m
7	R. Sterling	0.579	7	6	5	€ 120m
8	P. Pogba	0.549	55	9	28	€ 80m
9	H. Kane	0.545	4	140	6	€ 150m
10	S. Heung-Min	0.539	19	36	17	€ 50m

Results / Open Data

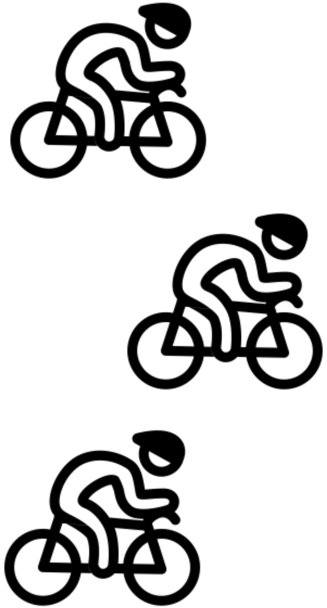
A Learn-to-Rank Approach for Predicting Road Cycling Race Outcomes

Leonid Kholkine, Thomas Servotte, Arie-Willem de Leeuw
Tom De Schepper, Peter Hellinckx, Tim Verdonck, Steven Latré

Predicting Sports Outcomes

- Mostly in team sports
  - Football, NBL, NBA, NHL, ...
- A range of techniques
  - Probabilistic models
  - Monte Carlo
  - Deep learning
  - …
- Classification problem
- External conditions are similar or easy to model

Predicting Cycling Outcomes



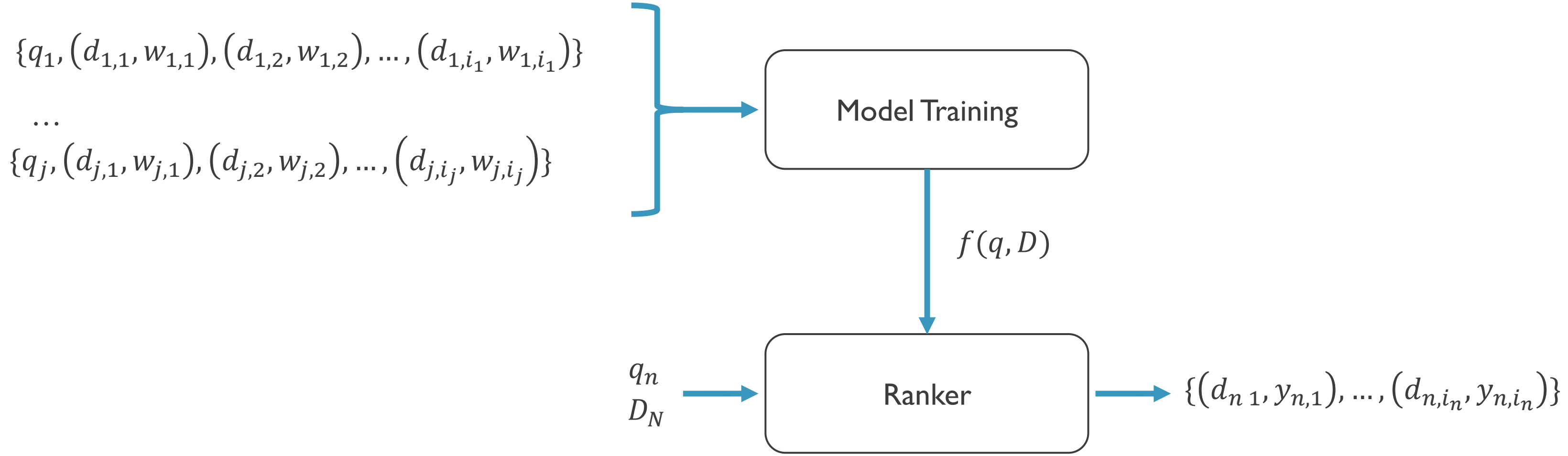
- It is a multi-contender sport
- Conditions change
  - Weather
  - Track
- Prone to several unpredictable events
  - Crashes
  - Injuries
- It is a "strategical endurance" sport
  - Rider will not go faster than need to
  - Not all riders are riding to win

State of the Art

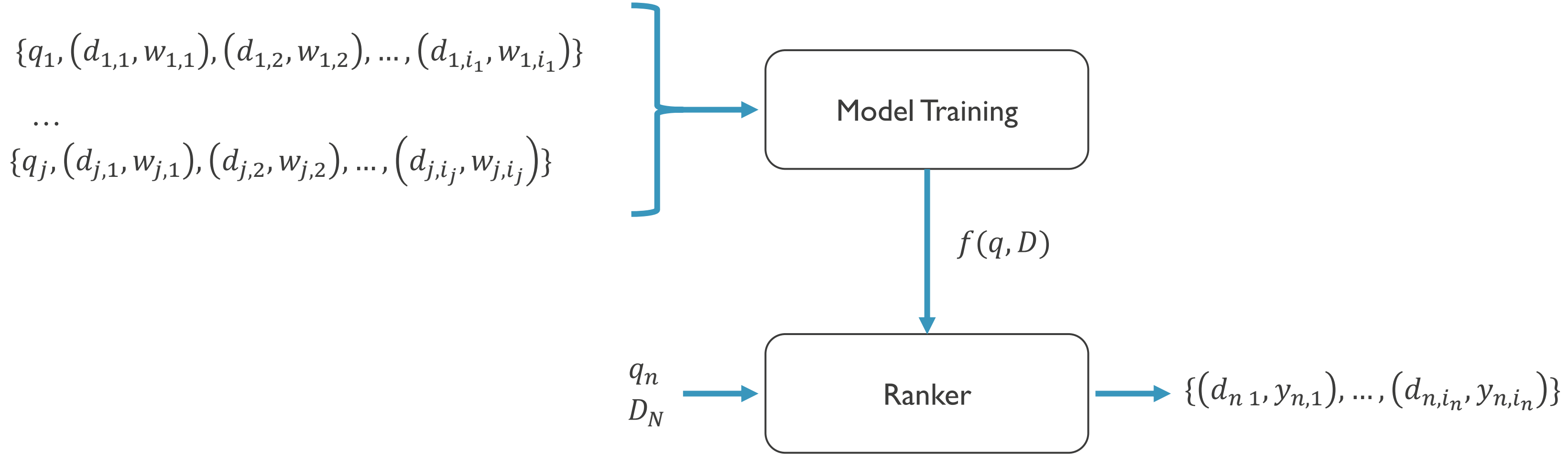
- Many classification models for 1 vs 1
- Some regression models for multi-contender sports
- None that optimize for ranking

Can a cycling race top 10 be predicted using a Learn-to-Rank approach?

Learn-to-Rank



Learn-to-Rank



Sets: Past editions of a race
Document: Rider
Query: Null
Weights: 20 for 1st, 19 for 2nd, etc…

Features

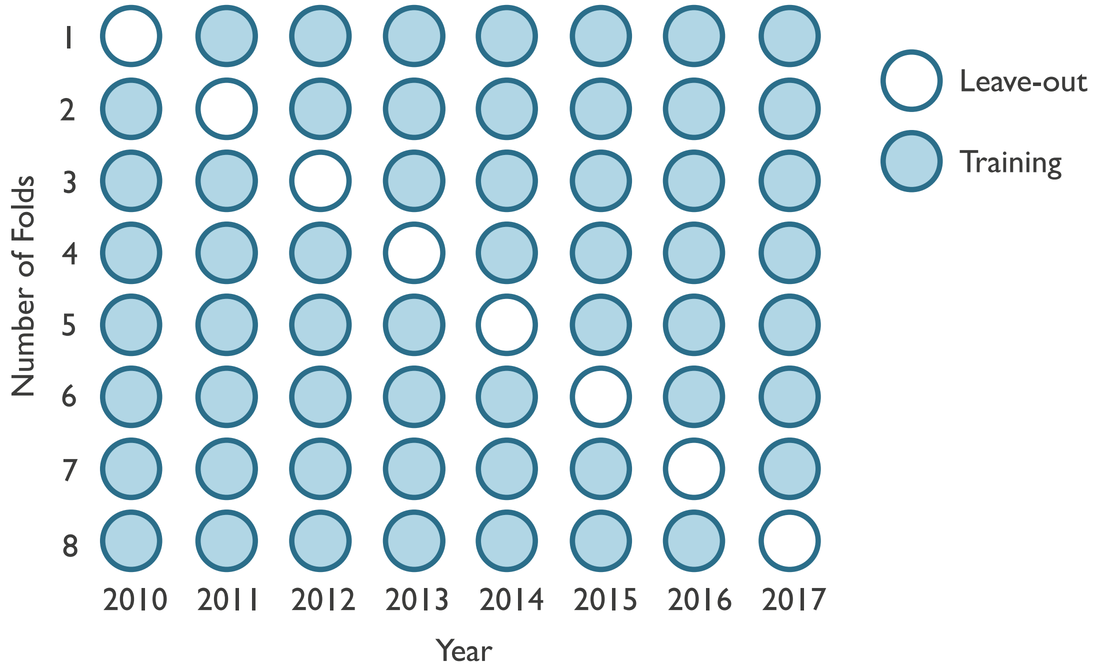
- Results from related races
- Overall performance in the past 3 years
- Evolution in the past 3 years
- Form
- Best result in the race
- Profile

Summary

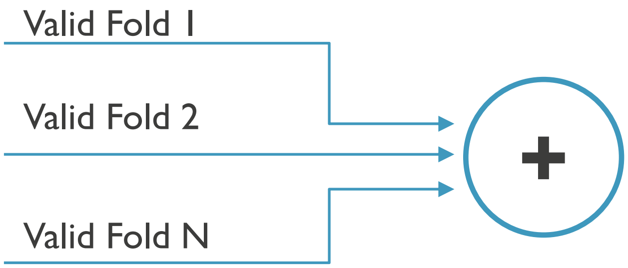
- Input: List of 20 editions of a race
- Output: Ranking of a race

- Problem
  - Not enough data

Yearly Cross Validation



Fold Ensemble



- Valid fold: Minimum number of iterations
- Valid ensemble: Minimum number of valid folds

Results

Race	Year	Fan NDCG	Model NDCG	Model Correct	Difference between Model and Fans
E3 Saxo Bank Classic	2018	0.58	0.54	6	0
E3 Saxo Bank Classic	2019	0.50	0.54	5	0
Ghent-Wevelgem	2018	0.68	0.62	6	-2
Ghent-Wevelgem	2019	0.23	0.32	3	0
Tour of Flanders	2018	0.62	0.67	6	-1
Tour of Flanders	2019	0.27	0.21	4	-1
Paris-Roubaix	2018	0.77	0.74	6	0
Paris-Roubaix	2019	0.35	0.44	4	0
La Flèche Wallonne	2018	0.57	0.60	5	2
La Flèche Wallonne	2019	0.55	0.61	5	1
Liège-Bastogne-Liège	2018	0.28	0.38	5	1
Liège-Bastogne-Liège	2019	0.43	0.31	3	-1
E3 Saxo Bank Classic	2021	0.32	0.37	3	-1
Ghent-Wevelgem	2021	0.41	0.63	5	3
Tour of Flanders	2021	0.69	0.69	7	0
La Flèche Wallonne	2021	0.84	0.76	6	-1
Liège-Bastogne-Liège	2021	0.69	0.81	8	1
Average:		0.52	0.55	5.12	0.12

Impact?

- Input to coaches for who to watch out for?
- An adapted model, can help with scouting
- Fan engagement

JULIAN ALAPHILIPPE

Deceuninck - Quick Step

VOLG

1

BELANGRIJKSTE RESULTATEN

- Tour de la Provence: 2
- Strade Bianche: 2
- Amstel Gold Race: 6
- La Flèche Wallonne: 1

2

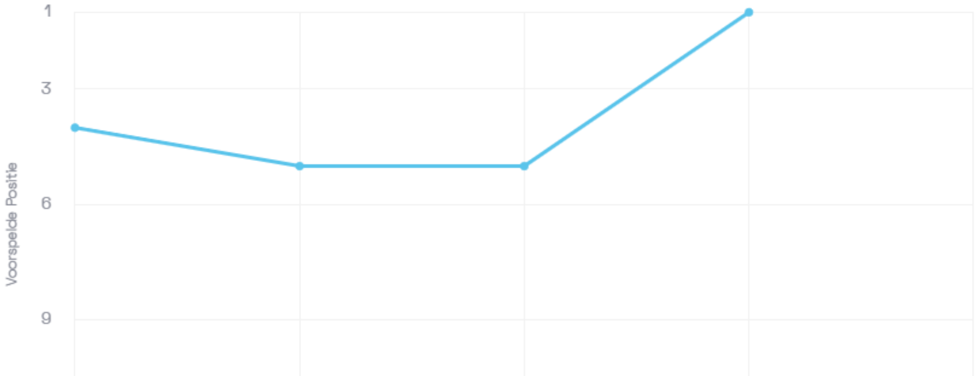
INFO

VERJAARDAG	VAN
1992-06-11	France

GEGEVENS

LENGTE	GEWICHT
1.73m 5ft 8in	62kgs 124lbs

HOE DE VOORSPELLING IN DE LOOP VAN DE TIJD VERANDERDE



Wie wint de Ronde? Artificiële intelligentie geeft het antwoord

01/04/21 om 07:51   Bijgewerkt om 13:46

Robben Scheire
Medewerker van Sport/Voetbalmagazine.

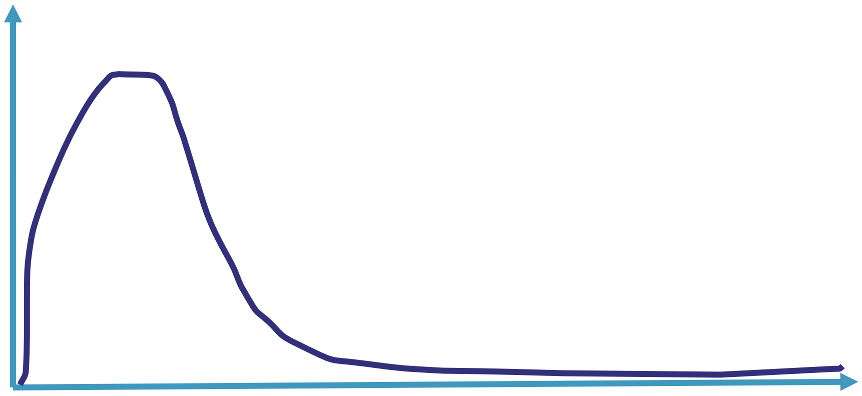
Onderzoekers van ID-lab, een Imec-onderzoeksgroep aan de Universiteit van Antwerpen, maken een website waarop ze via Artificial Intelligence (AI) de koers kunnen voorspellen.

Mathieu Van der Poel wint Ronde van Vlaanderen ... volgens onderzoekers UAntwerpen

ANTWERPEN De Nederlander Mathieu van der Poel wint zondag de Ronde van Vlaanderen. Dat voorspelt een nieuw ontwikkeld computersysteem aan de hand van artificiële intelligentie (AI). Op basis van historische prestatiedata tracht het systeem, dat door onderzoekers van imec en de Universiteit Antwerpen ontworpen is, de koersuitslagen van eendagswedstrijden te voorspellen. De voorspellingen zijn vanaf donderdag te raadplegen op wiewintdekoers.be.

Physiological Sensors

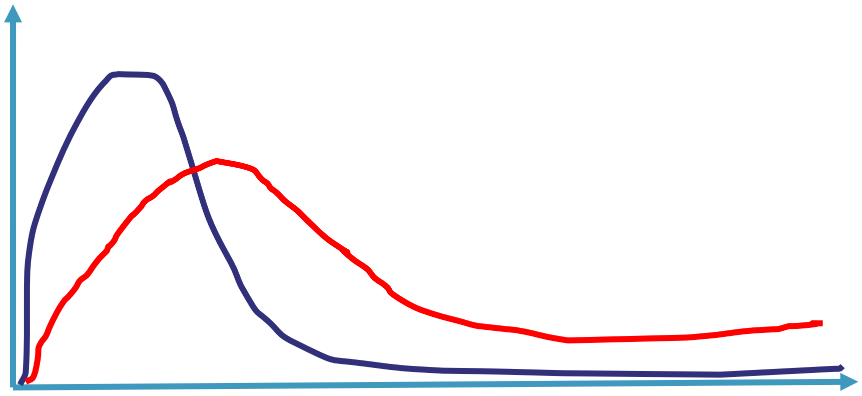
Anaerobic vs Aerobic Systems



ATP-CP

More in-depth explanation: https://www.youtube.com/watch?v=8Y_FdjI2v4I

Anaerobic vs Aerobic Systems

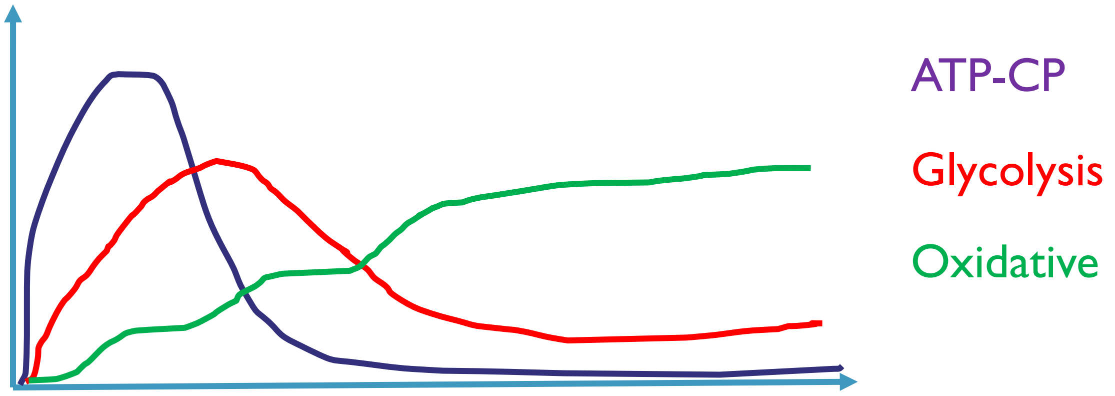


ATP-CP

Glycolysis

More in-depth explanation: https://www.youtube.com/watch?v=8Y_FdjI2v4I

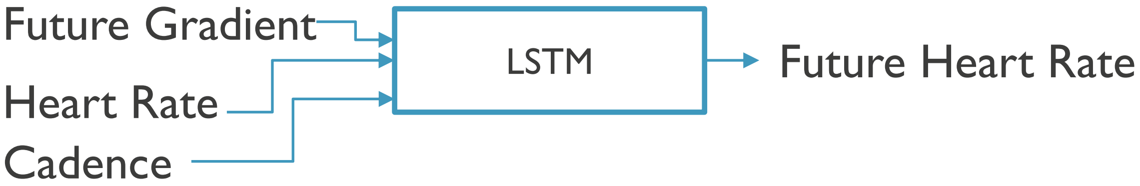
Anaerobic vs Aerobic Systems

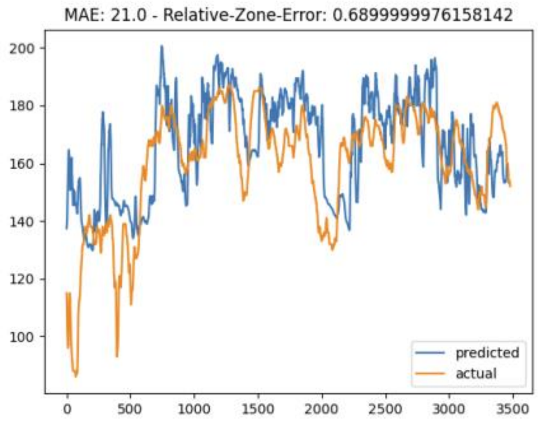


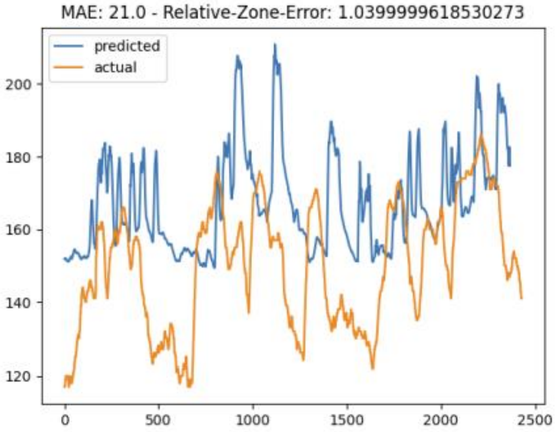
More in-depth explanation: https://www.youtube.com/watch?v=8Y_FdjI2v4I

Using LSTM Networks and Future Gradient Values to Forecast Heart Rate in Biking

Henry Gilbert e al.







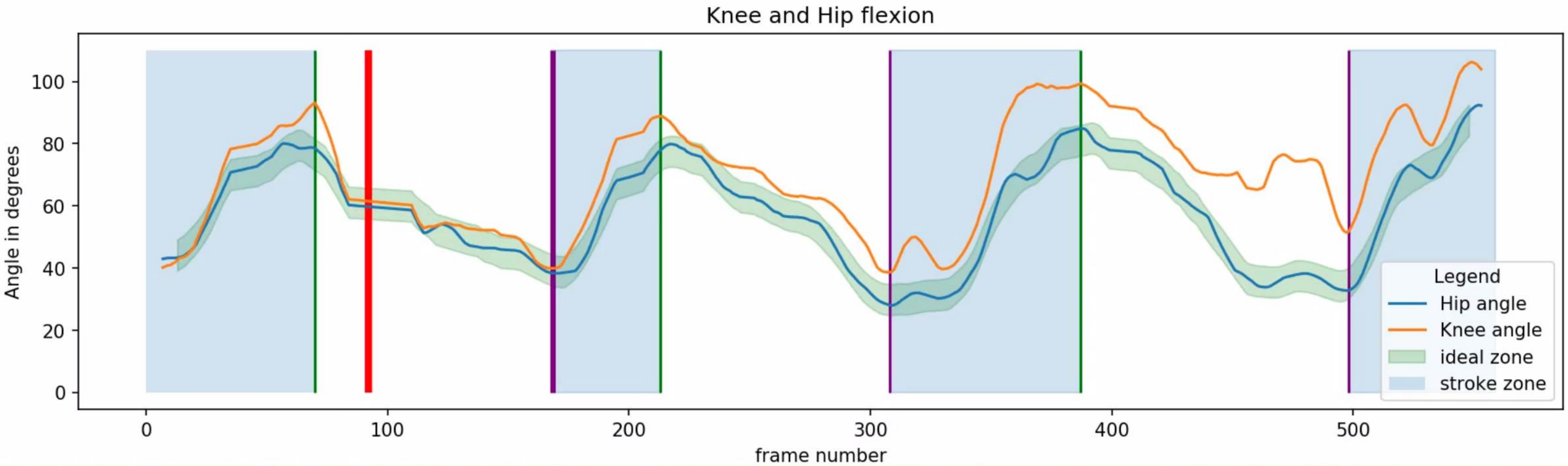
Using LSTM Networks and Future Gradient Values to Forecast Heart Rate in Biking

Henry Gilbert e al.

- Limited to 1 participant
- Only considering Heart Rate and Cadence
- Does not take into consideration power and power that will be used on climb

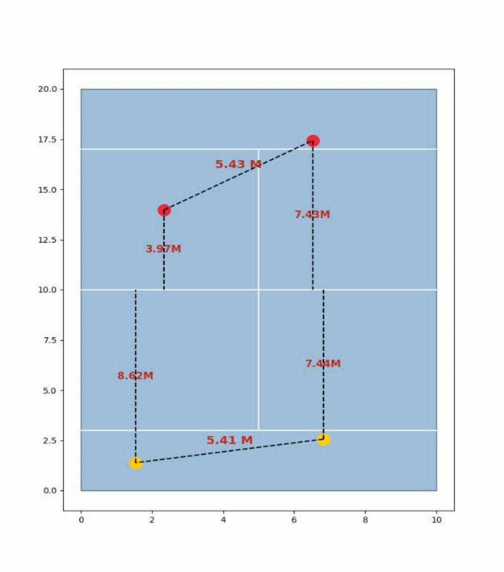
- Potential impact: Predicting how much a climb can impact the heart rate

Other Examples


Knee and Hip flexion
Angle in degrees
frame number
0
20
40
60
80
100
0
100
200
300
400
500
Legend
Hip angle
Knee angle
ideal zone
stroke zone


BEL

INTERPLAYER DISTANCE



What? The distance between two players of the same team

Why? An increased player gap increases team vulnerability

imec IDLab UNIVERSITEIT GENT Universiteit Antwerpen

A Strategic Framework for Optimal Decisions in Football 1-vs-1 Shot-Taking Situations: An Integrated Approach of Machine Learning, Theory-Based Modeling, and Game Theory

Calvin C. K. Yeung[1] and Keisuke Fujii[1,2,3*]

[1]Graduate School of Informatics, Nagoya University, Nagoya, Japan.
[2]Center for Advanced Intelligence Project, RIKEN, Osaka, Japan.
[3]PRESTO, Japan Science and Technology Agency, Saitama, Japan.

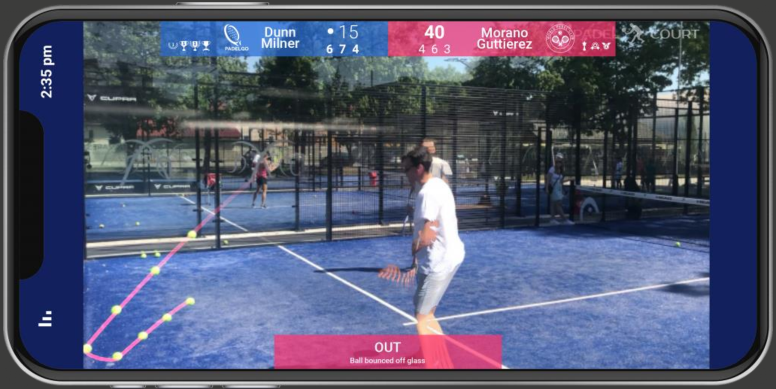
Commercial Solutions

Introducing

Opta Vision

PADELCOURT


2:35 pm
Run to the net
Avoid no
man's land


2:35 pm
Dunn
Milner
15
6 7 4
40
4 6 3
Morano
Guttierez
OUT
Ball bounced off glass


TRACKMAN

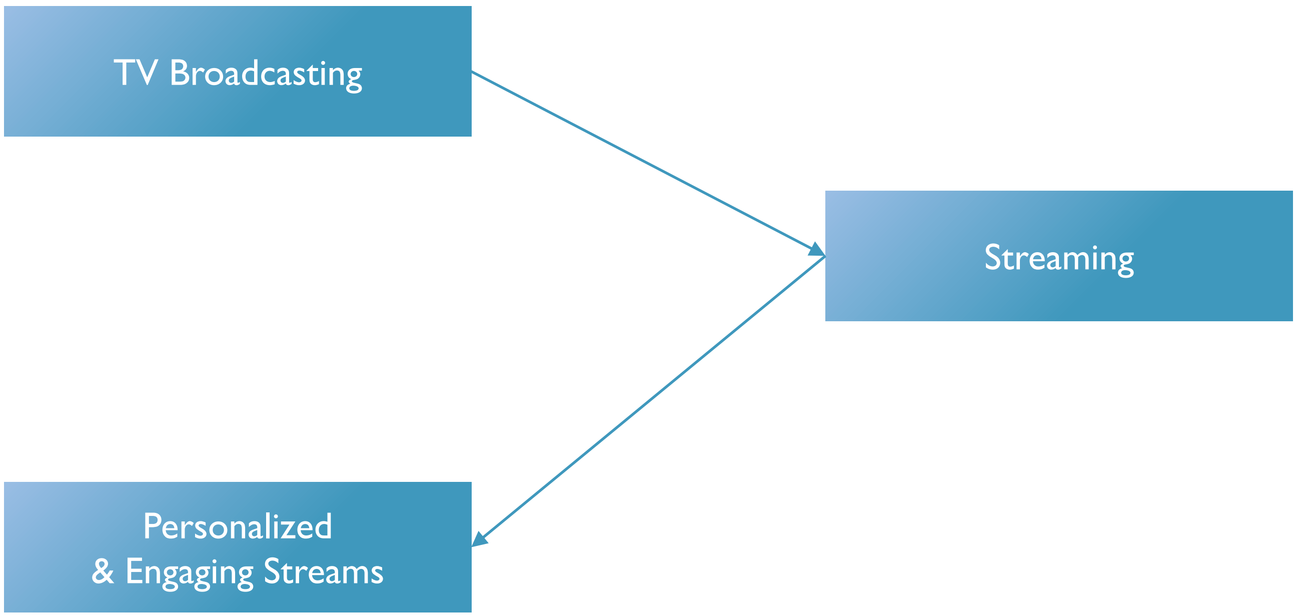

Impact
Dynamic Loft 15,0°
Spin Rate 1804
Attack Angle 3,1°
Spin Loft 11,9°
Club Path -0,7°L
Face To Path 1,0°R
Spin Axis 0,4°R
Face Angle 0,3°R
CLUB SPEED 116,8
BALL SPEED 173,8
SMASH FAC. 1,49
ATTACK ANG. 3,1
CARRY 293,1
TOTAL 316,6
SPIN RATE 1804
LAUNCH ANG. 13,9
CLUB PATH -0,7
FACE ANG. 0,3


DIGITAL ATHLETE MODEL
MITIGATION MODEL
1.9 %
01

What the future (could) hold

1. Sport Sustainability & Growth

Personalized Fan Experience



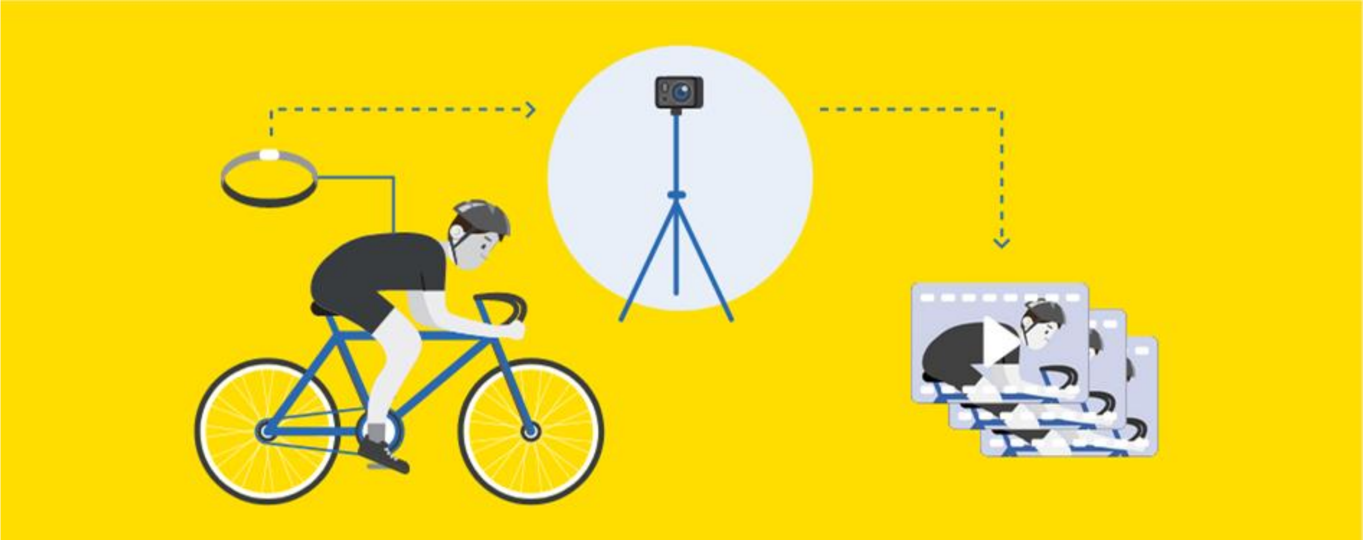
Fully Automatic Camera for Personalized Highlight Generation in Sporting Events

Robbe Decorte and Steven Verstockt
(IDLab UGent - imec)

BOSA2024

2024-02-06

STRADA ecosystem



Hardware:

- Tripod
- Smart camera
- Sensor data capturing device
- Computing unit

Software:

- Collecting all data streams
- Automatic video clip generation
- Performance
- Fan engagement

Automated Match Summary

Event Data → Semantic Event Data → Summary

Manuel Labeling

LLM

JSON

2:41 Player A tackled player Z receiving a red card

Summary Text

https://www.raccoons.be/resources/insights/the-generative-ai-playground-part-1:-automated-soccer-match-report

VENTOUX:

Video ENrichment To Optimize User eXperience of cycling race broadcasts

Jelle Vanhaeverbeke and Steven Verstockt
(IDLab UGent - imec)

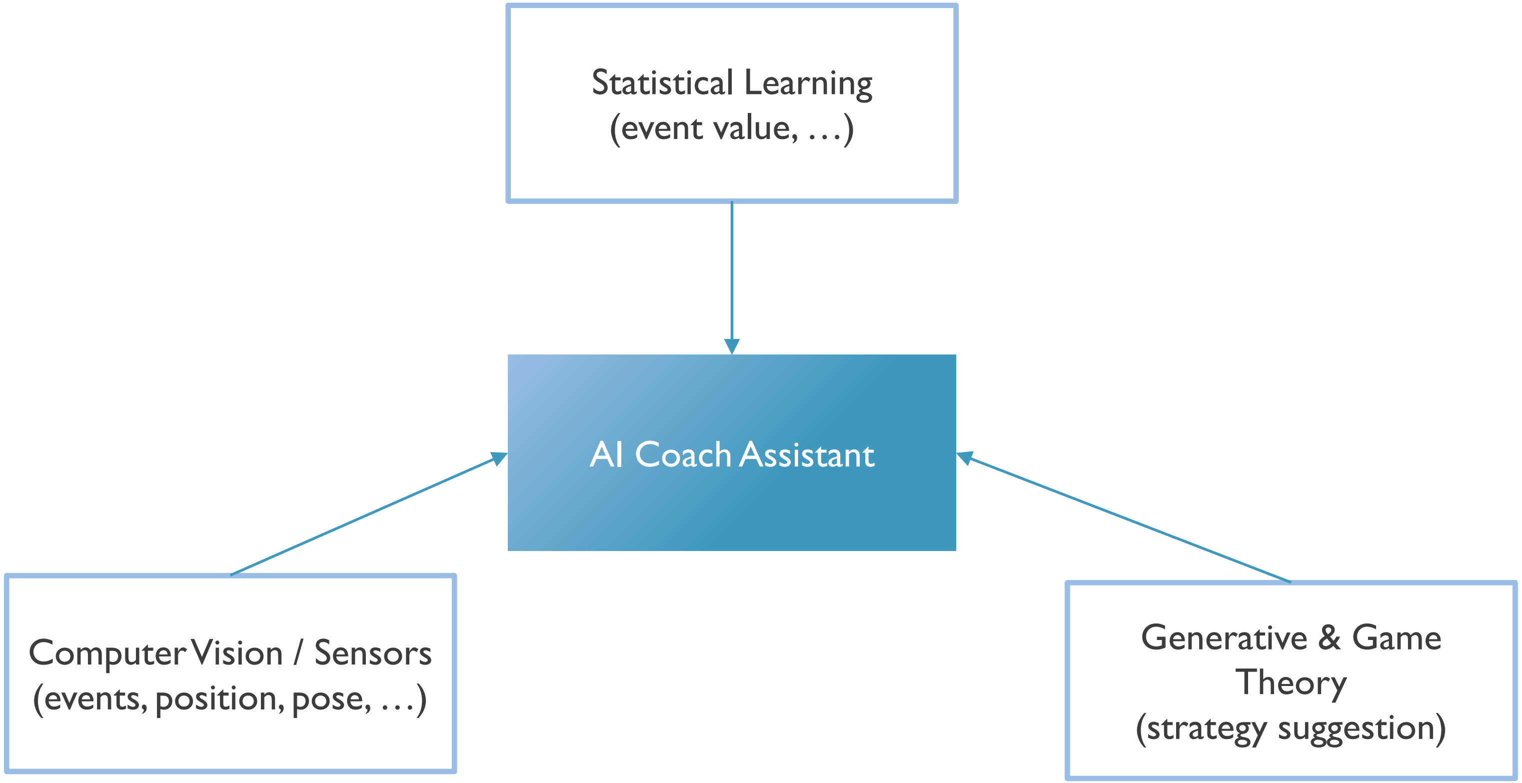
BOSA2024

2024-02-06

Demo: Omloop het Nieuwsblad 2023

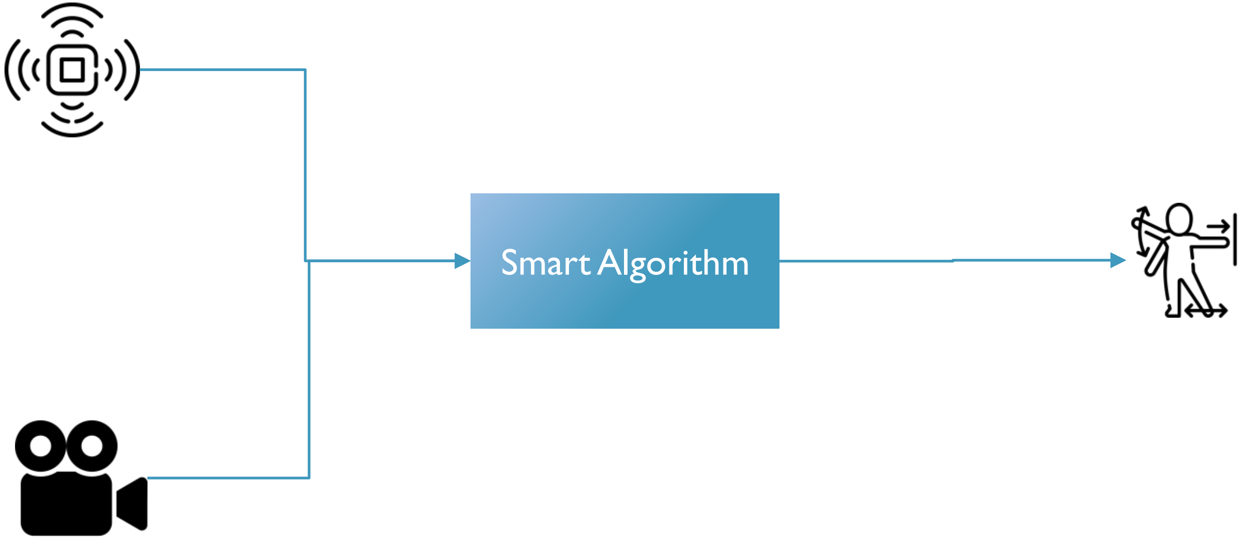
2. AI Coach Assistant

Team Coach Assitant



3. Automated Biomechnical Feedback

Algorithms to personalize



4. Fitness coach assistant & Personalized approach to cardiovascular training

Correlation properties of heart rate variability to assess the first ventilatory threshold and fatigue in runners

Bas Van Hooren, Bram Mennen, Thomas Gronwald, Bart C. Bongers & Bruce Rogers

https://doi.org/10.1080/02640414.2023.2277034

Correlation properties of heart rate variability to assess the first ventilatory threshold and fatigue in runners

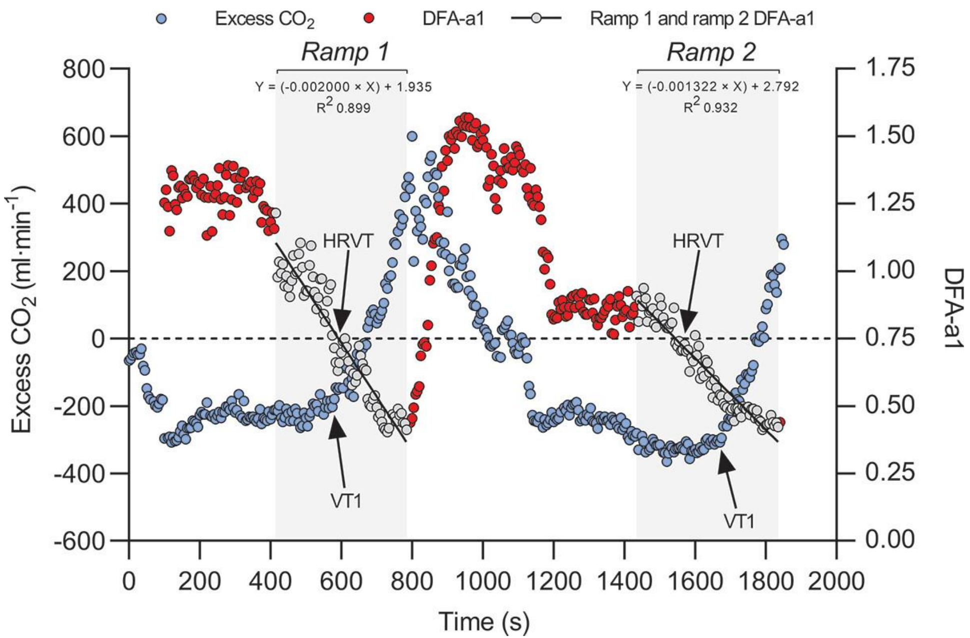
Bas Van Hooren, Bram Mennen, Thomas Gronwald, Bart C. Bongers & Bruce Rogers

Twenty-nine participants **(24 males, 5 females; mean ± SD age 24.7 ± 5.6 years; body mass 75.1 ± 11.2 kg; body height 178.1 ± 7.9 cm)**, that were free of any moderate (for previous 3 months) or minor (for previous 1 month) musculoskeletal injuries, were not taking any medication that could influence running performance, were aged 18–45 years, comfortable with treadmill running, and had a body mass index (BMI) <30 kg/m^2 volunteered to participate in this study.

https://doi.org/10.1080/02640414.2023.2277034

Correlation properties of heart rate variability to assess the first ventilatory threshold and fatigue in runners

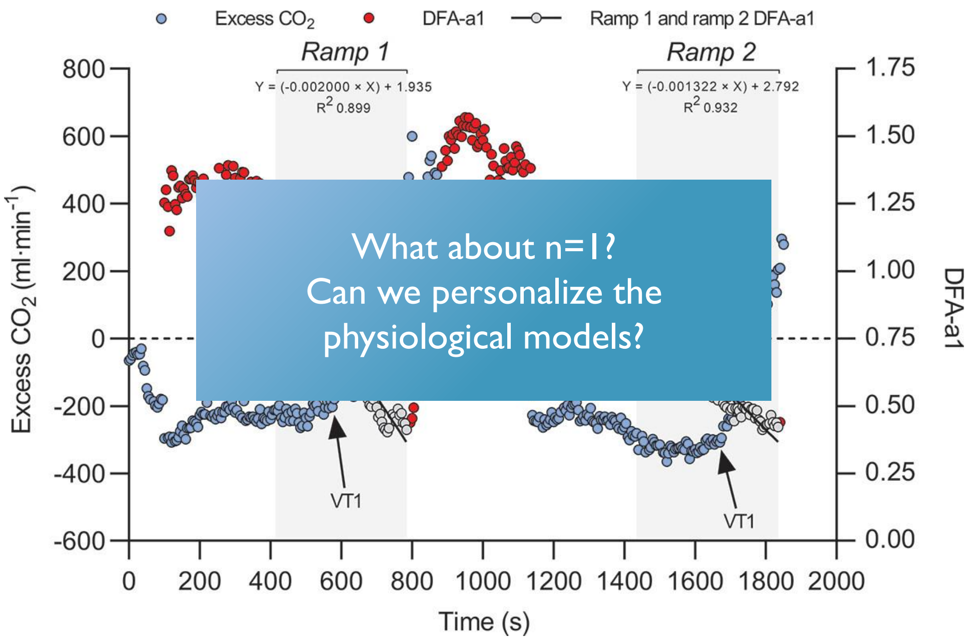
Bas Van Hooren, Bram Mennen, Thomas Gronwald, Bart C. Bongers & Bruce Rogers



https://doi.org/10.1080/02640414.2023.2277034

Correlation properties of heart rate variability to assess the first ventilatory threshold and fatigue in runners

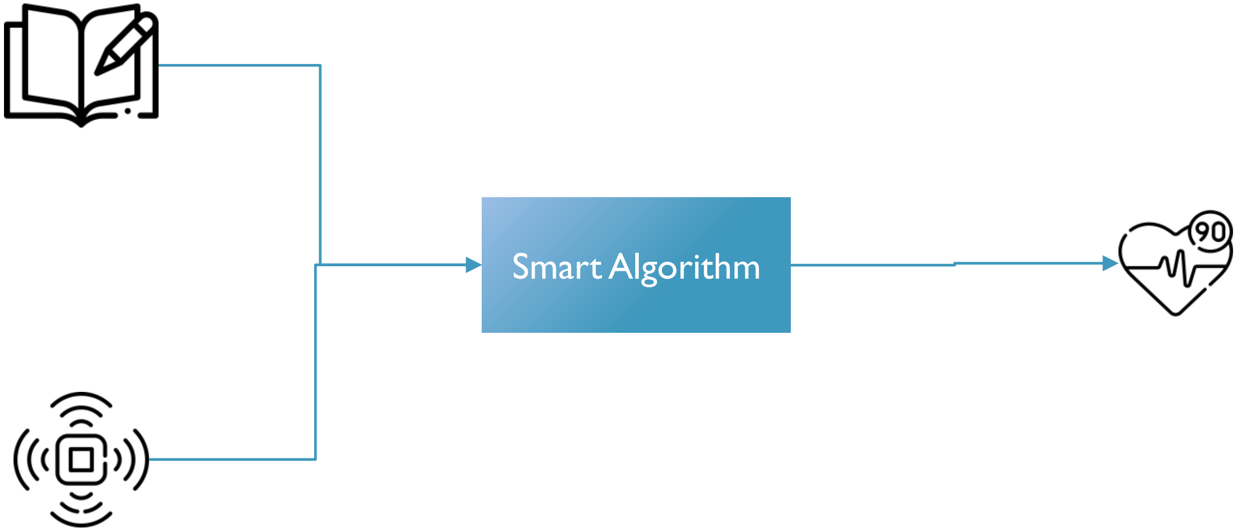
Bas Van Hooren, Bram Mennen, Thomas Gronwald, Bart C. Bongers & Bruce Rogers



What about n=1?
Can we personalize the physiological models?

https://doi.org/10.1080/02640414.2023.2277034

Algorithms to personalize



What are the challenges?

Challenges & Risks

- Explainability

Comment on "Black Box Prediction Methods in Sports Medicine Deserve a Red Card for Reckless Practice: A Change of Tactics is Needed to Advance Athlete Care"

Jakim Berndsen[1*] and Derek McHugh[1†]

[1*]Kitman Labs, Dublin, Ireland.

https://arxiv.org/pdf/2204.02402.pdf

Challenges & Risks

- Explainability
- Privacy & regulation

Professional sports may run afoul of privacy law in collecting player biometric data

Co-authored by Matthew Hennessy, Partner, Peter Divitcos, Associate, Malcolm Liu, Solicitor and Chloe Tsatsos, Law Graduate.

https://www.dentons.com/en/about-dentons/news-events-and-awards/news/2023/october/professional-sports-may-run-afoul-of-privacy-law-in-collecting-player-biometric-data

Challenges & Risks

- Explainability
- Privacy & regulation
- Gap data science <> sports science

REVIEW PAPER **Open Access**

Machine learning methods in sport injury prediction and prevention: a systematic review

Hans Van Eetvelde[1*], Luciana D. Mendonça[2,3,4], Christophe Ley[1], Romain Seil[5] and Thomas Tischer[6]

https://sportsdatascience.be/BOSA2024/

Challenges & Risks

- Explainability
- Privacy & regulation
- Gap data science <> sports science
- Resources
  - Gap will become bigger with sports / teams that have money

Challenges & Risks

- Explainability
- Privacy & regulation
- Gap data science <> sports science
- Resources
  - Gap will become bigger with sports / teams that have money
- “The more money you poor, the more medals you get”

“AI enables & enhances technology in sports”

Questions?

https://bit.ly/connect_to_leon

leonid.kholkine@uantwerpen.be