



PASSING TACTICAL RELEVANCE DURING SOCCER MATCHES DETERMINED BY MACHINE LEARNING

Merlin, M¹, Moura, F², Torres, R¹, Principe, V³, Cunha, S.A¹

1: State University of Campinas, 2: State University of Londrina, 3: Rio de Janeiro State University

SPORT AND MATHEMATICS



What is the importance of sport for mathematics?



http://www.rtv.unicamp.br/?video_listing=educacao-fisica-e-matematica

grant# 2016/50250-1 - 2017/20945-0 - 2018/19007-9
São Paulo Research Foundation (FAPESP)

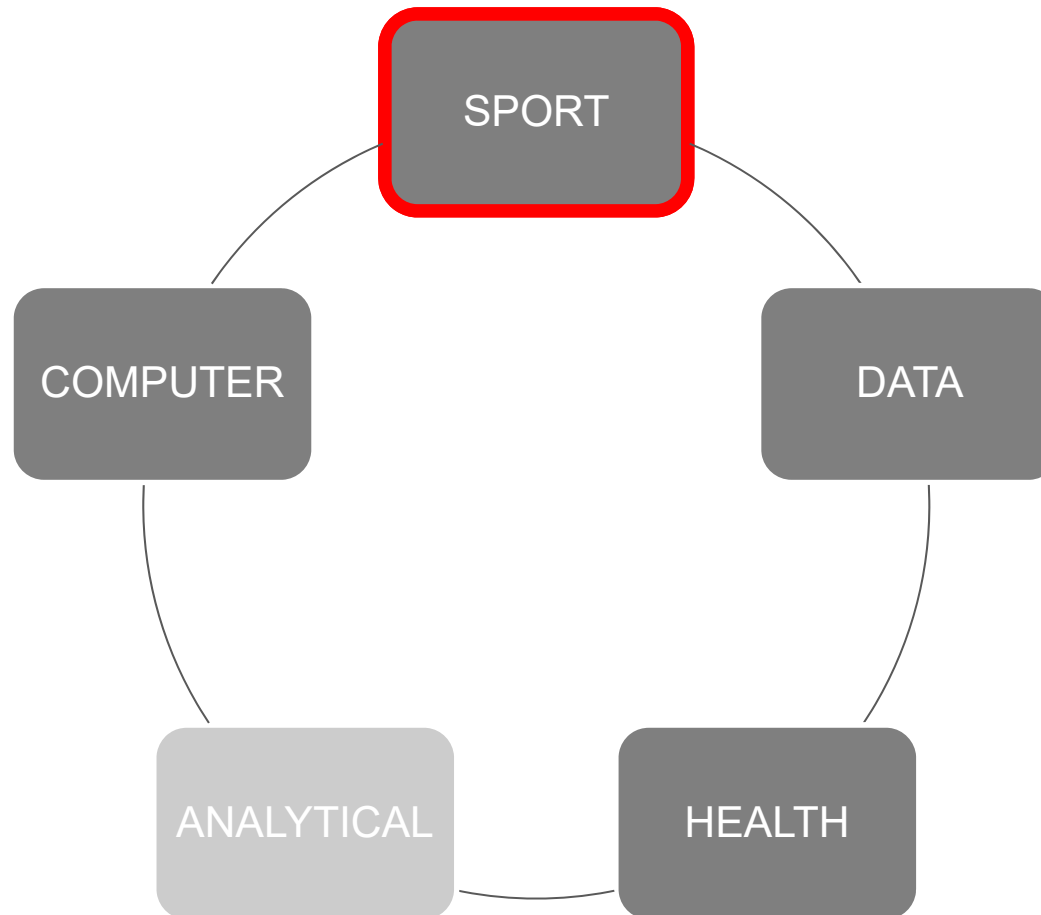
SPORT AND MATHEMATICS

SPORT



MATHS

SPORT AND COMPUTER



SOCCKER

ACCURACY

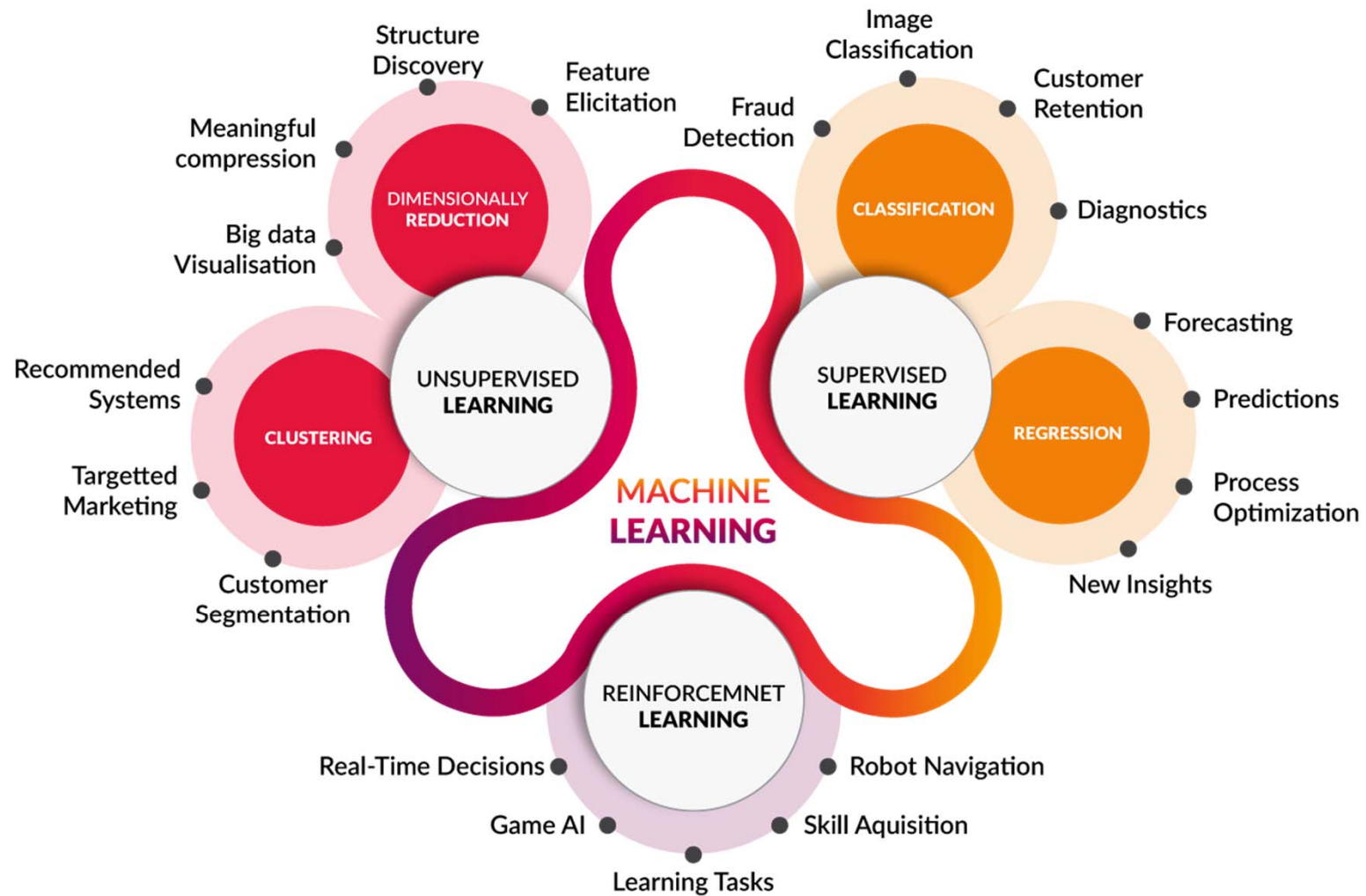


BALL



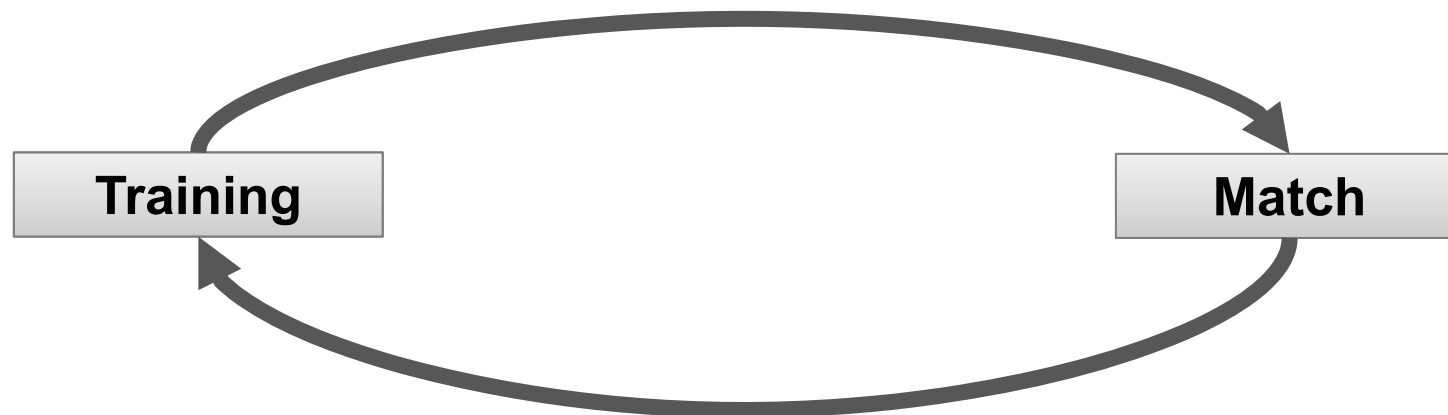
PHYSICAL-TACTICAL-TECHNICAL

MACHINE LEARNING



SOCCER

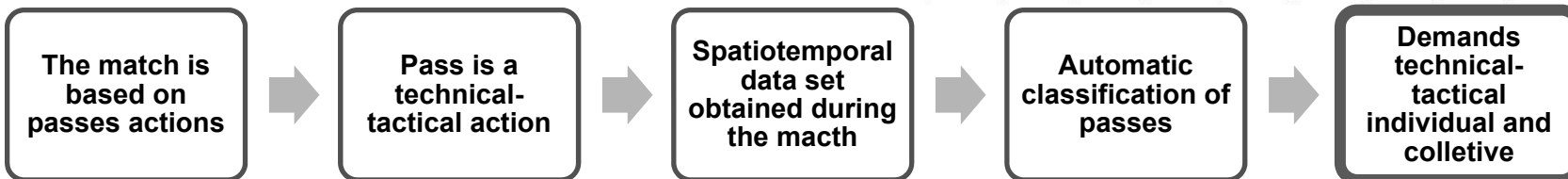
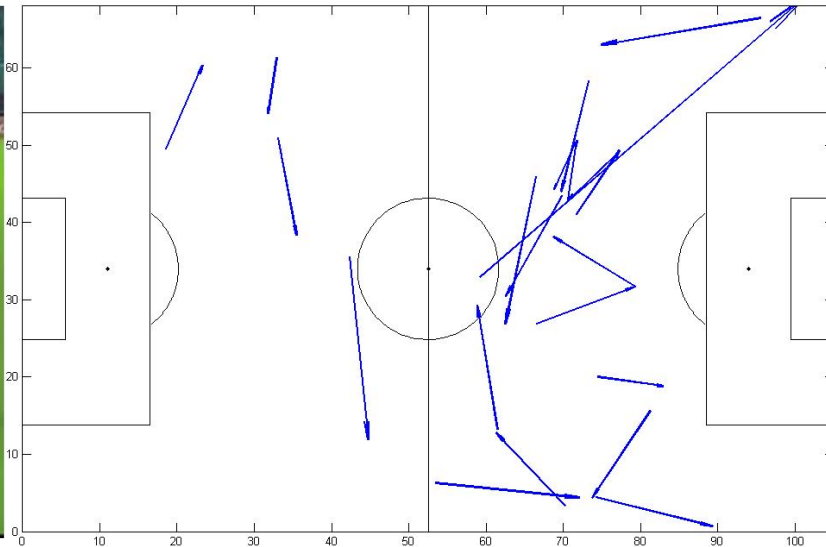
- Intermittent team sports.
- Machine Learning techniques used in soccer matches.
- Training to Match – Match to Training.



SOCCKER PASSES

- Passes are the most used technical-tactical action during the match and has been considered an important performance indicator.
- Usually, the pass is analyzed based on effectiveness (successful or not successful) limiting the quality of information.

SOCCER PASSES



- ✓ understanding of the game;
- ✓ understanding the specificities of positions;
- ✓ planning, intervention and training control;
- ✓ detection of talents.

SOCCKER PASSES

- GRAPH VISUAL RHYTHMS (Rodrigues *et al.*, 2019): successful tactical strategies used in soccer matches (passes between players).
- Classifying passes based on most significant concept to performance would allow more relevant analysis about the match.
- Mathematics can help to answer the importance of passes for soccer without human subjectivity

SOCCKER PASSES

- Passing TACTICAL RELEVANCE (TR)
- Machine learning to classify a soccer problem (passes complexity - very bad to very good)
- Tactical advantage provided by the passing in a given real match context.

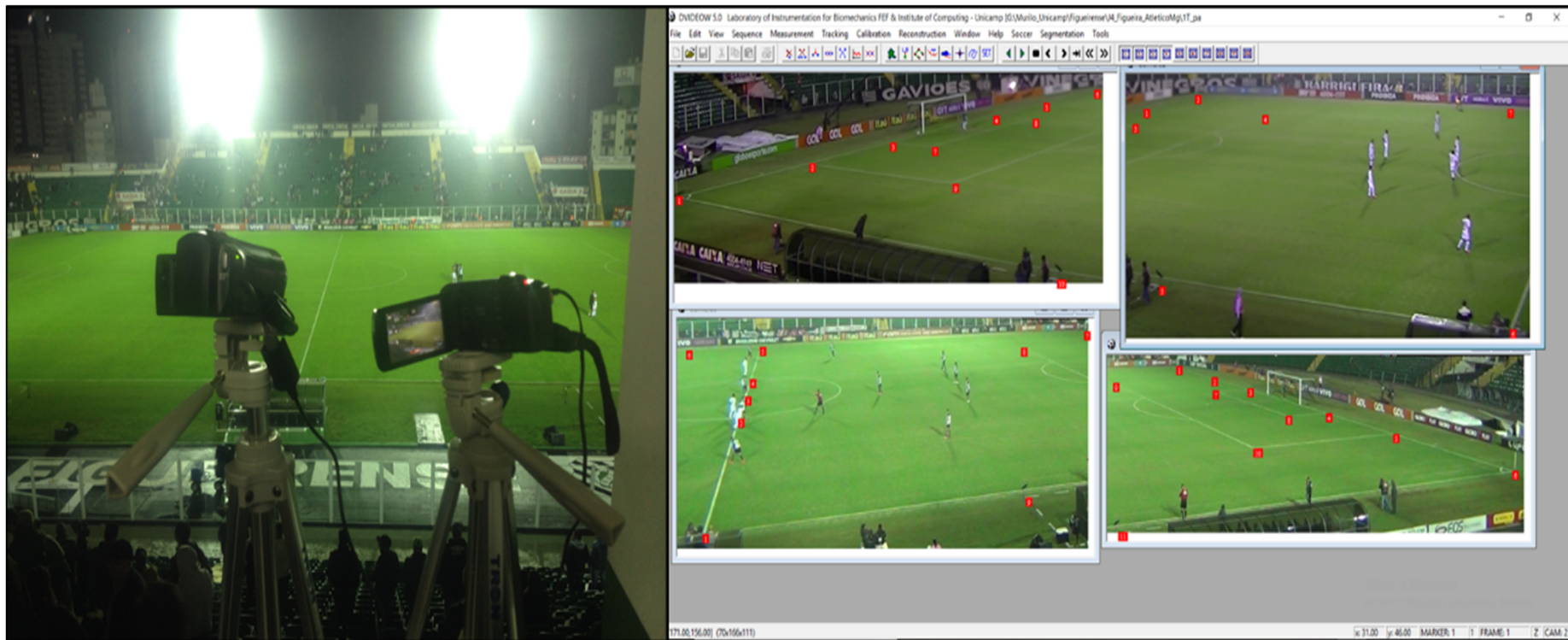
AIM

- Propose an automatic pass classification model in soccer matches based on tactical relevance concept, from spatiotemporal data.

METHODOLOGY

- 4 official matches of Professional Brazilian Football Championship 2016.
- Tracking by Dvideo 15 Hz (0.3 m - BARROS *et al.* 2007).
- Smoothing by digital filter *Butterworth* (Matlab[®]2017).
- Scout by DVideo interface.

METHODOLOGY



grant# 2016/50250-1 - 2017/20945-0 - 2018/19007-9
São Paulo Research Foundation (FAPESP)

METHODOLOGY

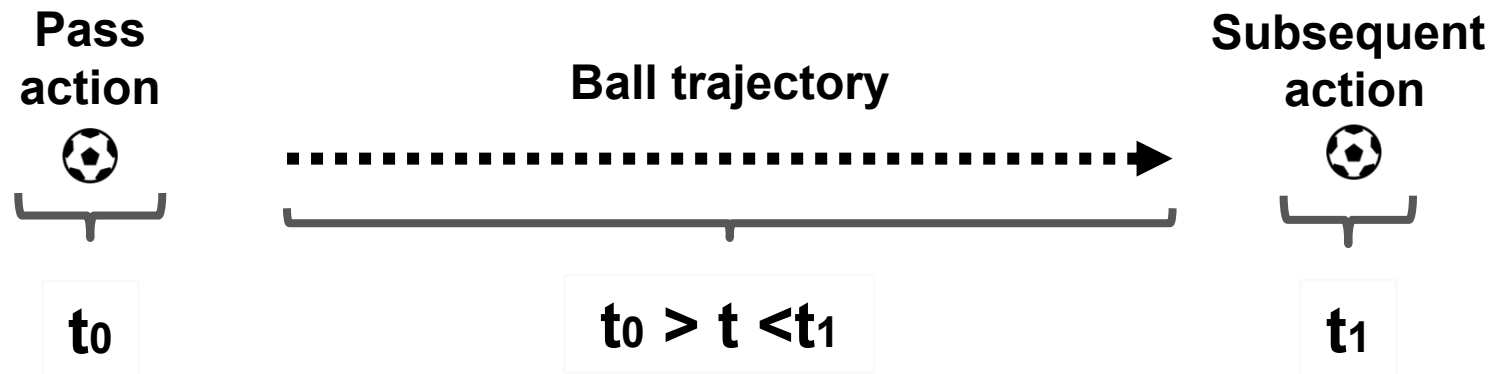
- Supervised dataset framework:

Passing predictive variables.

Passing response variables (labeling process).

METHODOLOGY

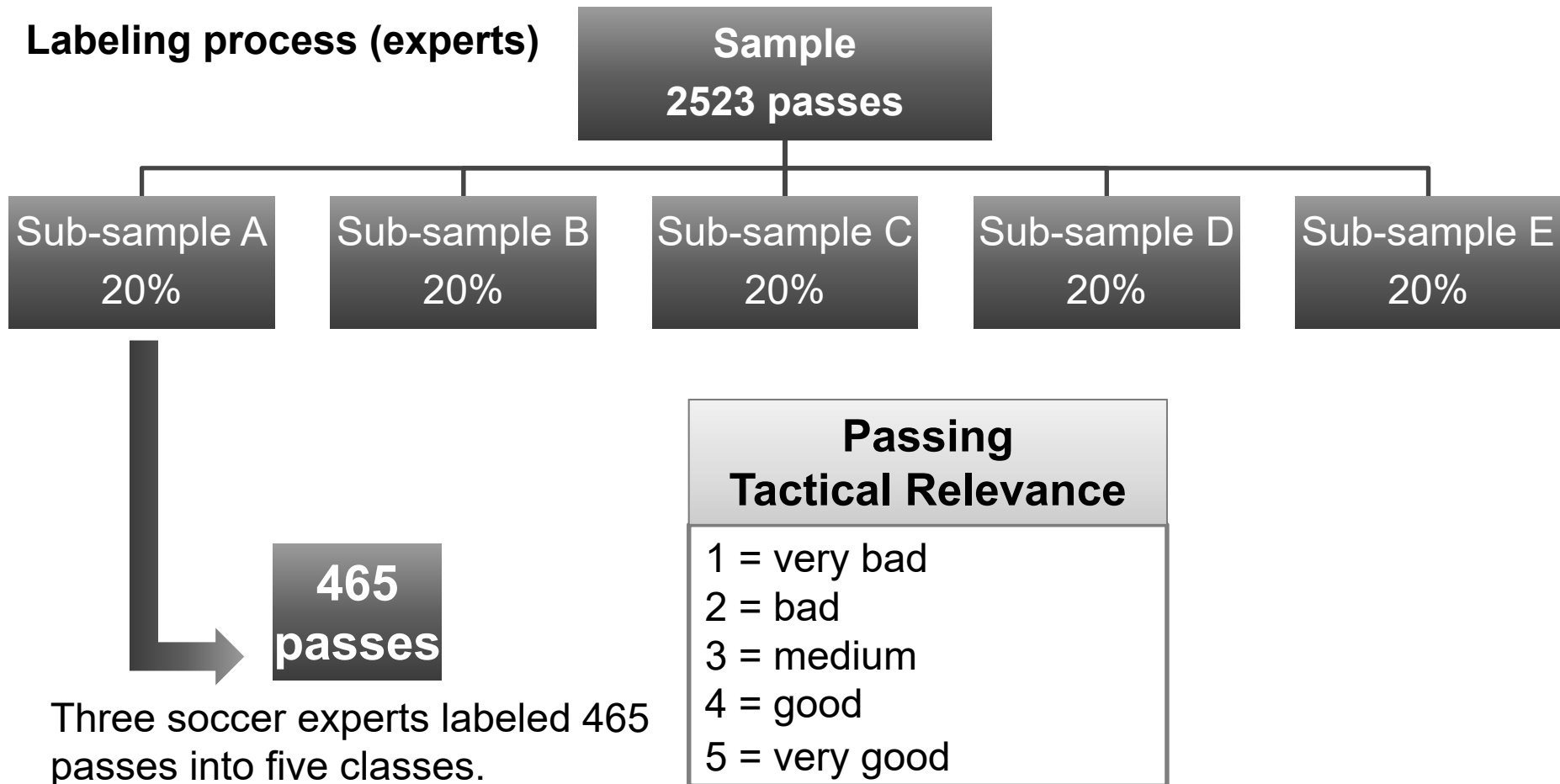
Passing predictive variables



- Density
- Nearest opponent
- Outplayed opponent
- Distance, angle
- Opponent btw receiver and target
- Key pass
- Assistance, etc.

33 predictive variables

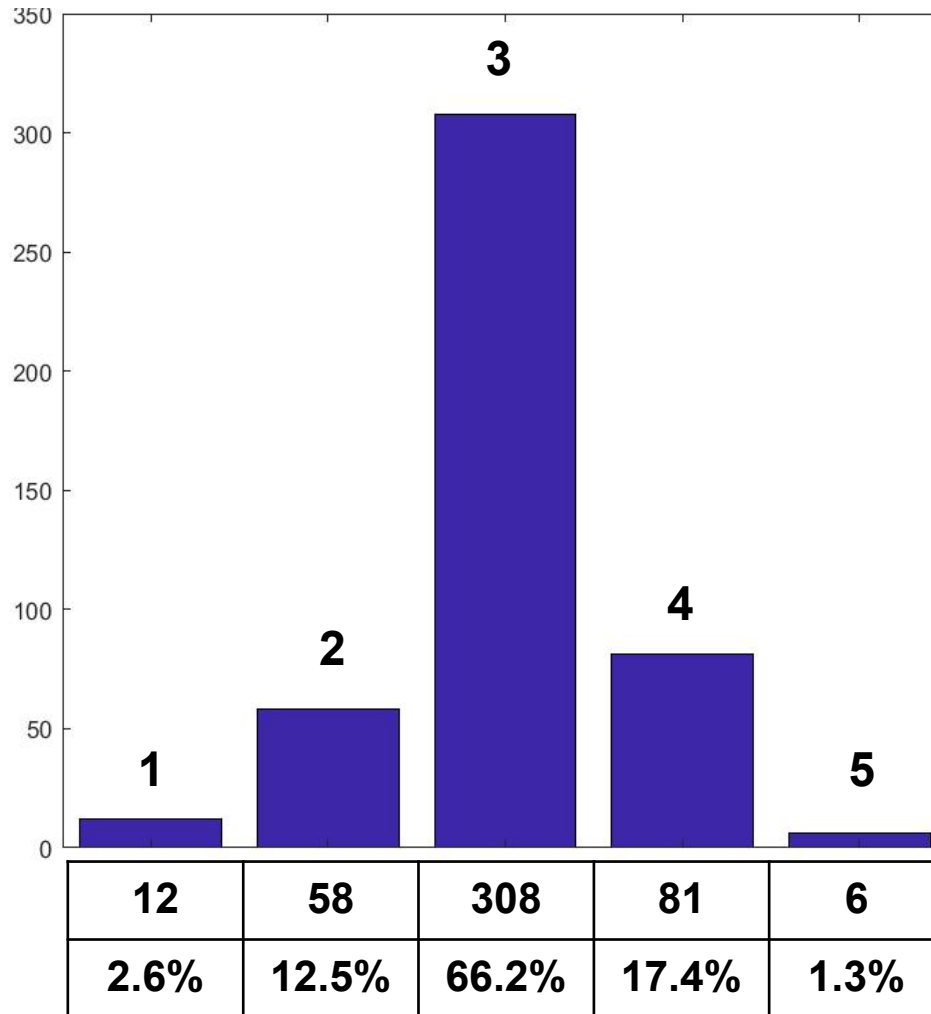
METHODOLOGY



METHODOLOGY

- Classification process
- Software Python 3.7 - scikit-learn v0.20.3 library;
- Random Forest Classifier (RF);
- Split stratified K-fold ($n = 3$);
- Normalized and balanced accuracy.

RESULTS



Labeling process (experts)

**Passing
Tactical Relevance**

1 = very bad
 2 = bad
 3 = medium
 4 = good
 5 = very good

RESULTS

Labeling process (experts)

	Very Bad	Bad	Medium	Good	Very Good	Total
Very Bad	7	0	0	0	0	7
Bad	9	46	8	1	0	64
Medium	0	5	262	56	0	323
Good	0	0	5	59	2	66
Very Good	0	0	0	0	5	5

- ✓ Agreement = 81,5%
- ✓ *Kappa* = 0,65 (substancial)

RESULTS

- Passing Tactical Relevance:

Normalized accuracy (K-fold) = 0.80 ± 0.01

Balanced accuracy (K-fold) = 0.50 ± 0.04

RESULTS

- accuracy = 80.0% (passes classification - 5 classes) - similar Horton & Gudmundsson (2014) accuracy = 85.0% (passes quality classification - 3 classes).

FINAL CONSIDERATIONS

- Passing Tactical Relevance: multivariate and nonlinear problem.
- Predictor variables based on the opinion of experts, and a classification model based on a nonlinear algorithm.

FINAL CONSIDERATIONS

- Test the classifier with three classes.
- Increase number of labeled samples from the more extreme classes.
- Test other classifiers, parameters and performance evaluation of the classifiers.

TAKE HOME MESSAGE

- Practical applications: player ranking, offensive sequence of actions (time series) and training planning.
- Data to Players: how can we do this?
- Quality of Data – Relevance of Analysis.
- Talent identification and development.



scunha@unicamp.br

**grant# 2016/50250-1 - 2017/20945-0 - 2018/19007-9
São Paulo Research Foundation (FAPESP).
CAPES, Finance Code 001.**