Interpretable Prediction and Large-Scale Analysis of Judging in Professional Boxing

Mason duBoef, Thomas Romeas, Mathieu Charbonneau, and Allan Svejstrup

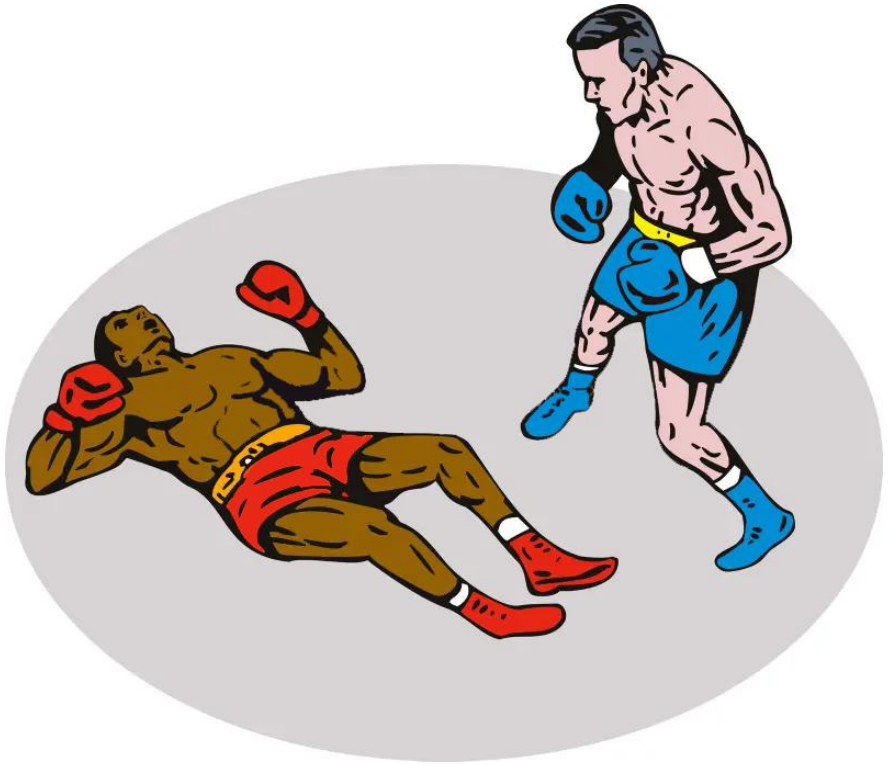
JABBR

INS QUÉBEC

The Sport of Boxing

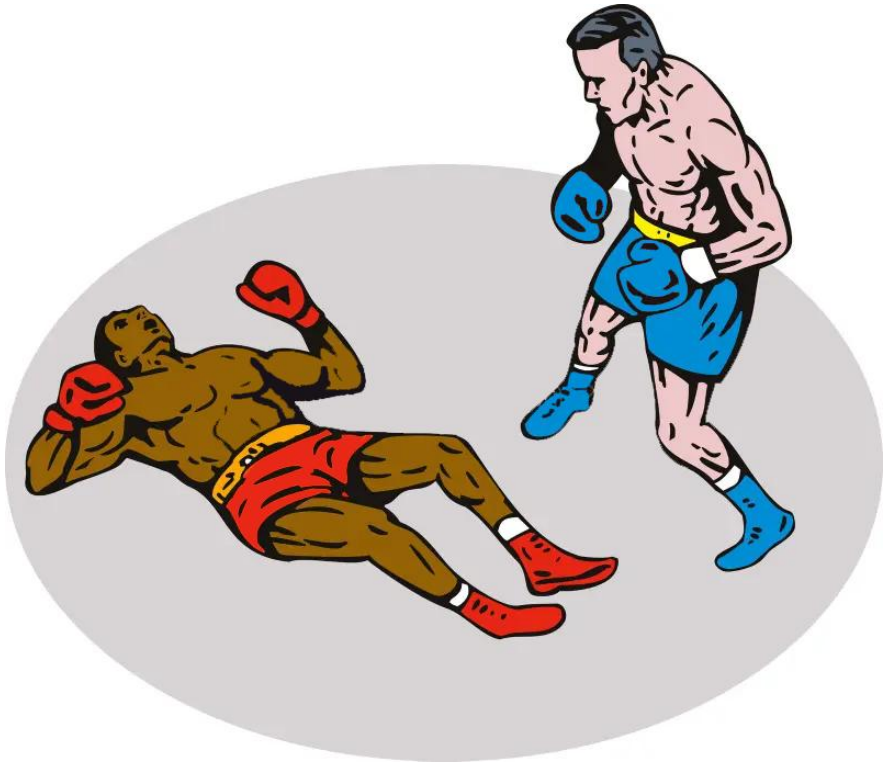
The Sport of Boxing

1. Knockout your opponent



The Sport of Boxing

1. Knockout your opponent



2. Match ends and is decided by judges

46.5% of matches

Scoring is subjective and intransparent

Round	Judge 1	Judge 2	Judge 3
1			
2			
3			

Existing Data Limitations

- Clicker-based punch stats
  - Inaccurate
  - Lack of detail

Existing Data Limitations

- Clicker-based punch stats
  - Inaccurate
  - Lack of detail
- Manual annotation
  - Limits sample size to about 50 rounds


JABBR
INS QUÉBEC
Landed 5 Thrown 19
Time Type Status Quality
02:32 R Straight Head Missed
02:32 L Hook Body Missed
02:33 R Uppercut Head Landed
02:34 L Straight Head Landed
02:35 R Uppercut Head Landed
02:35 L Hook Head Missed
02:38 R Uppercut Head Missed
02:39 L Hook Head Missed
Landed 5 Thrown 13
Time Type Status Quality
02:28 L Straight Head Landed
02:29 L Overhand Head Missed
02:34 L Straight Head Missed
02:35 L Uppercut Head Missed
02:36 L Straight Head Missed
02:36 L Uppercut Head Missed
02:37 L Straight Head Landed
02:42 L Straight Head Landed
Balance
Stance
Distance
12 2:28
jabbr.ai

Goals

(1) Build models to accurately predict judges' scorecards

(2) Identify what factors are most important to judges

Data Set

1,003 bouts

7,323 rounds

Detailed end-of-round statistics

Round-by-round scores

Mapping Methods

Neural Network

- Multi-layer perceptron (MLP)

Mapping Methods

Neural Network

- Multi-layer perceptron (MLP)

Points-Based System (PB)

Specific performance metric

$$R_{\text{points}} = aR_1 + bR_2 + cR_3 + \dots$$

$$B_{\text{points}} = aB_1 + bB_2 + cB_3 + \dots$$

Mapping Methods

Neural Network

- Multi-layer perceptron (MLP)

Points-Based System (PB)

Specific performance metric

$$R_{\text{points}} = aR_1 + bR_2 + cR_3 + \dots$$
$$B_{\text{points}} = aB_1 + bB_2 + cB_3 + \dots$$

Weight assigned to specific metric

Mapping Methods

Neural Network

- Multi-layer perceptron (MLP)

Points-Based System (PB)

- Optimized with gradient descent

Specific performance metric

$$R_{\text{points}} = aR_1 + bR_2 + cR_3 + \dots$$
$$B_{\text{points}} = aB_1 + bB_2 + cB_3 + \dots$$

Weight assigned to specific metric

Mapping Methods

Neural Network

- Multi-layer perceptron (MLP)

Points-Based System (PB)

- Optimized with gradient descent

Specific performance metric

$$R_{\text{points}} = aR_1 + bR_2 + cR_3 + \dots$$
$$B_{\text{points}} = aB_1 + bB_2 + cB_3 + \dots$$

Weight assigned to specific metric

Ratio of points →

$$R_\varphi = \frac{R_{\text{points}} + D}{B_{\text{points}} + D}$$

Predicted score →

$$R_\Theta = \frac{(R_\varphi)^S}{(R_\varphi)^S + 1}$$

Canelo Alvarez vs Terence Crawford

Judge Scorecard AI Prediction

Rnd	1	2	3	4	5	6	7	8	9	10	11	12
Red	50%	43%	50%	55%	62%	28%	44%	46%	16%	38%	4%	15%
Blue	50%	57%	50%	45%	38%	72%	56%	55%	84%	62%	96%	85%

(1) Prediction Accuracy

Model Accuracy

Measure of Predictive Accuracy	PB Model	MLP Model	Tiny PB Model
Pairwise Comparison Acc.	75.98%	75.52%	75.54%
Agreement with Majority	77.59%	77.31%	77.24%
Mean Squared Error	0.383	0.392	0.403

Model Accuracy

39 metrics

Measure of Predictive Accuracy	PB Model	MLP Model	Tiny PB Model
Pairwise Comparison Acc.	75.98%	75.52%	75.54%
Agreement with Majority	77.59%	77.31%	77.24%
Mean Squared Error	0.383	0.392	0.403

Model Accuracy

Measure of Predictive Accuracy	PB Model	MLP Model	Tiny PB Model
Pairwise Comparison Acc.	75.98%	75.52%	75.54%
Agreement with Majority	77.59%	77.31%	77.24%
Mean Squared Error	0.383	0.392	0.403

Model Accuracy

22nd percentile →

20th percentile →

Rank	Judge	Accuracy	Rounds
1	Judge A	98.33%	60
2	Judge B	97.83%	46
3	Judge C	96.51%	86
4	Judge D	95.45%	44
5	Judge E	94.44%	108
6	Judge F	93.75%	48
7	Judge G	93.75%	48
8	Judge H	93.55%	62
9	Judge I	93.55%	62
10	Judge J	93.48%	46
	...	...	...
177	Judge K	76.09%	23
178	Judge L	76.04%	48
—	*PB Model (Test Set)*	*75.98%*	*1450*
179	Judge M	75.86%	29
180	Judge N	75.77%	130
181	Judge O	75.61%	41
—	*Tiny PB Model (Test Set)*	*75.54%*	*1450*
—	*MLP Model (Test Set)*	*75.52%*	*1450*
182	Judge P	75.37%	67
183	Judge Q	75.00%	42
	...	...	...
225	Judge R	60.87%	23
226	Judge S	60.71%	28
227	Judge T	54.55%	22
Avg	**All Judges**	**81.41%**	**7323**

(2) Identifying What Matters To Judges

Not All Punches Are Equal

Not All Punches Are Equal

Measure of Predictive Accuracy	No Impact Differentiation	With Impact Differentiation
Pairwise Comparison Acc.	71.89%	73.15%
Agreement with Majority	73.10%	74.62%
Mean Squared Error	0.482	0.448

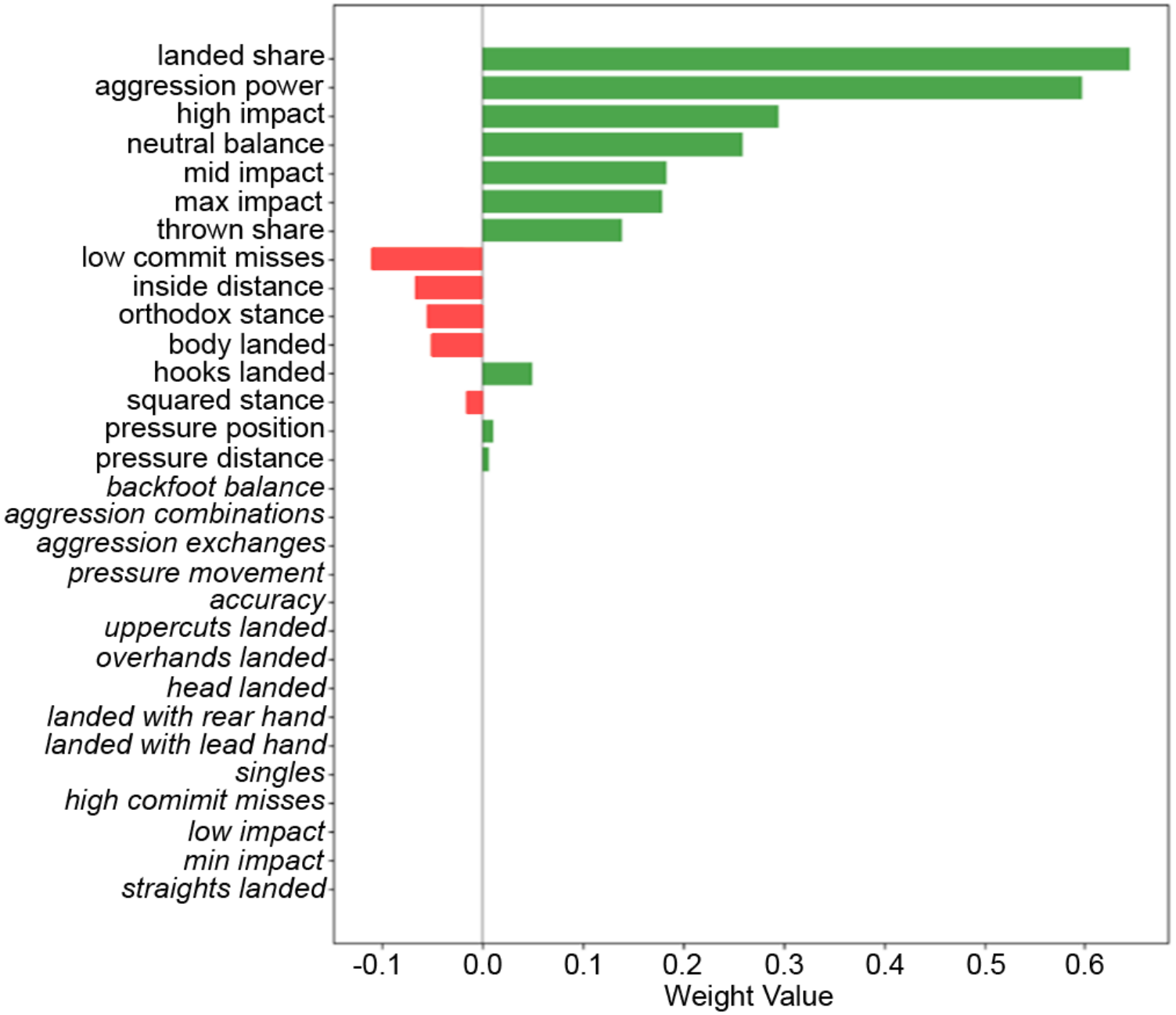
Not All Punches Are Equal

Measure of Predictive Accuracy	No Impact Differentiation	With Impact Differentiation
Pairwise Comparison Acc.	71.89%	73.15%
Agreement with Majority	73.10%	74.62%
Mean Squared Error	0.482	0.448

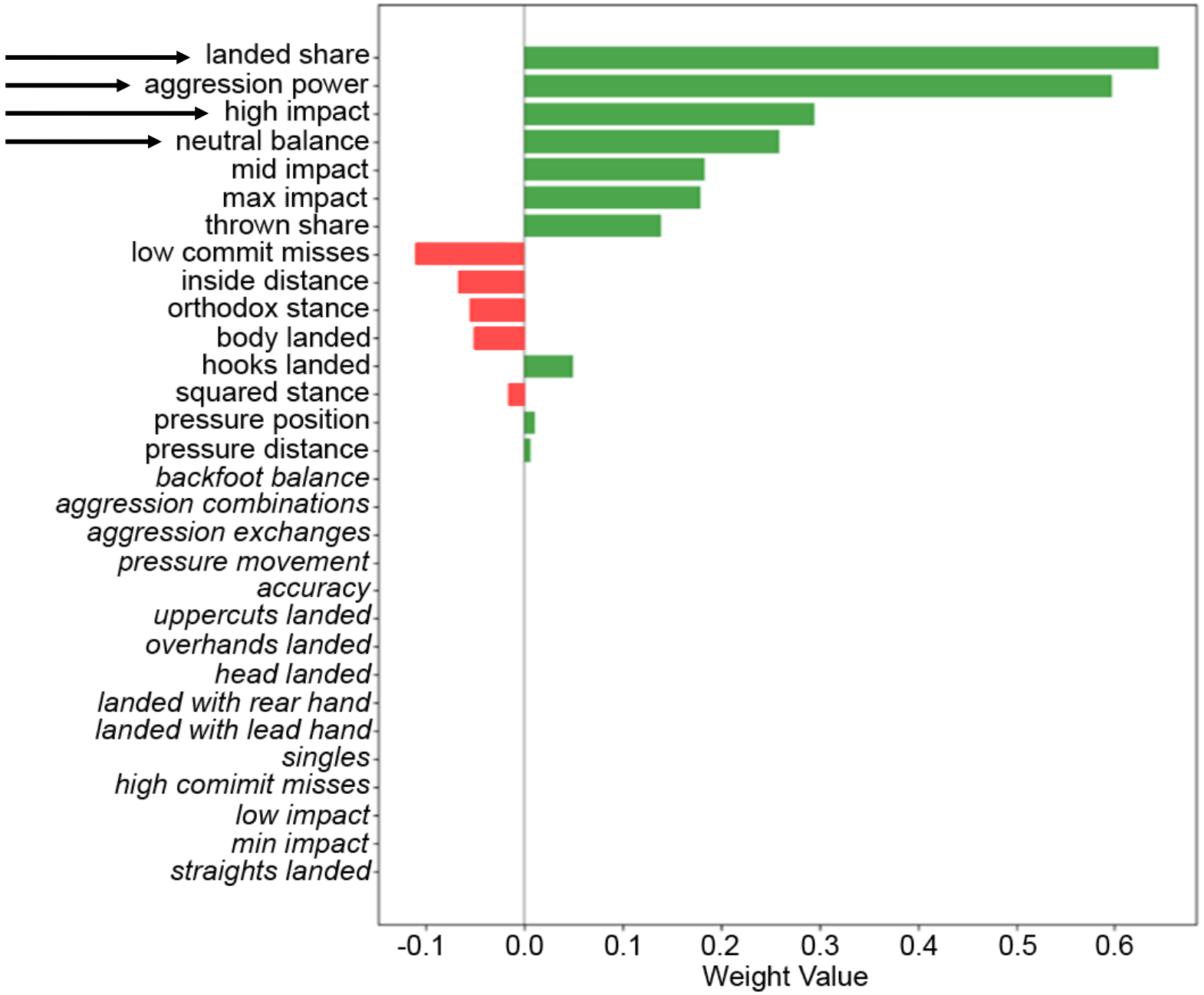
Metric	Normalized Weight
missed	0.24
min impact	1.00
low impact	1.45
mid impact	2.54
high impact	4.40
max impact	10.50

exponential

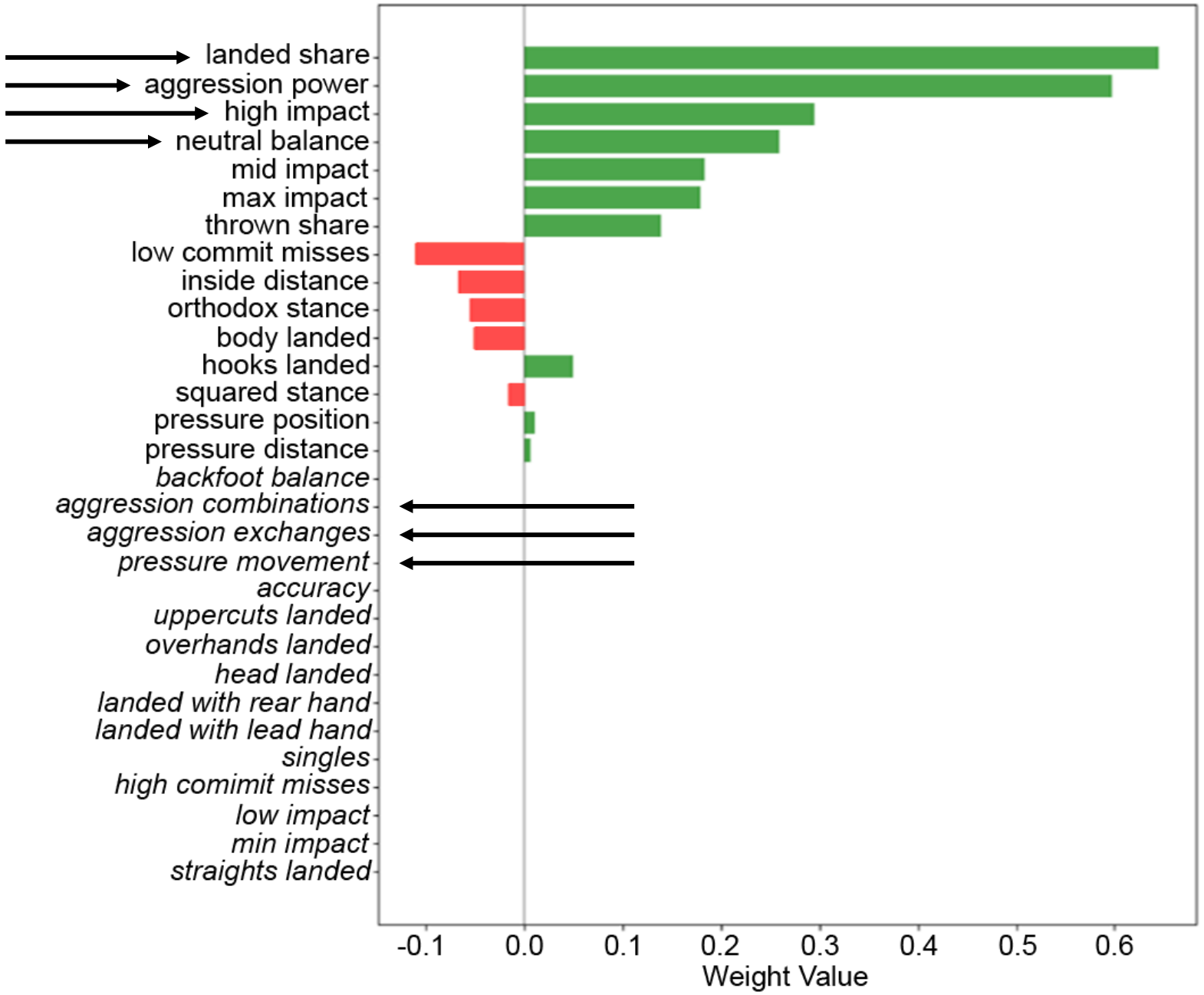
Feature Selection with L1 Logistic Regression



Feature Selection with L1 Logistic Regression



Feature Selection with L1 Logistic Regression



Limitations

- Missing contextual info
- Missing body language

- Tracking based on single camera dirty feed
- Outdated version of DeepStrike

Takeaways

Takeaways

(1) Simple points-based scoring achieves pro-level accuracy

- Consistent
- Transparent
- Unbiased
- Scalable

Takeaways

(1) Simple points-based scoring achieves pro-level accuracy

- Consistent
- Transparent
- Unbiased
- Scalable

(2) Punch impact and throwing with power drive decisions

Future Work

Future Work

- Address data limitations

Future Work

- Address data limitations
- Identify stylistic differences between judges

Future Work

- Address data limitations
- Identify stylistic differences between judges
- Investigate impact of biasing features
  - Fighter nationality, ranking, popularity

Future Work

- Address data limitations
- Identify stylistic differences between judges
- Investigate impact of biasing features
  - Fighter nationality, ranking, popularity
- Extend to other sports (MMA, fencing, etc.)

Thank You

Questions?

Mason duBoef
Jabbr, University of Massachusetts Amherst
mduboef@gmail.com

Special thank you to Prof. P.M. Aronow (Yale)