Benchmarking feature selection algorithms for optimal classification and dataset comprehension: a biomechanical application

J. I. M. Parmentier[a,b], F. M. Serra Bragança[a] and B. J. van der Zwaag[b]

[a]Clinical Sciences Department, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands; [b]Pervasive Systems Group, EEMCS, University of Twente, Enschede, The Netherlands

1. Introduction

With the increasing amount of data collected in both equine clinical and field settings, datasets are growing larger with high dimensionality and complexity. Machine learning algorithms are becoming a go-to solution to classify data but remain black-box solutions that offer little transparency to the end-users, such as veterinarians and trainers. To reduce the dimensionality and complexity of the datasets, many feature selection algorithms (FSAs) are available, but currently, there is no 'textbook' about which FSA to use for which application. In this work, we propose an FSA benchmarking method to optimally select features of a high-dimensional dataset for binary classification models. Moreover, our method includes a post-analysis of the features selected by each algorithm to understand better the dataset and the influence of the data provenance on the classification outcome. We illustrate this method with a surface classification ('Hard' or 'Soft') of equine locomotion data.

2. Methods

2.1. Data collection

A total of 96 horses were equipped with six inertial measurement units (IMUs) attached to the withers, sacrum, and the lateral aspects of the distal limbs (EquiMoves®), sampling at 200Hz. Each IMU node contains 3D high-g and low-g accelerometers (±16g and ±200g resp.), and a 3D gyroscope (±2000dps). The nodes were aligned with the body segments of the horse to accurately represent accelerations and rotation rates encountered in the dorsoventral/proximodistal, mediolateral and anteroposterior directions, as shown in Figure 1.

Data were collected in-hand at different gaits (walk, trot), in different directions (straight line, left circle, right circle) and on different surfaces (hard, soft).

Figure 1. Horse equipped with the 6 IMUs. Green arrow: *x*-axis; blue arrow *y*-axis; red arrow: *z*-axis.

2.2. Data processing and features extraction

The high- and low-g accelerometers signals were merged for optimal precision and range (Bosch et al. 2018). Signals expressed in the local IMU frame were automatically segmented into different gaits (Walk and Trot) (Serra Bragança et al. 2020) and direction of the movement (straight: 0; left circle: −1; right circle: 1), and then manually labelled (measurement notes) as Hard (bricks, concrete, undeformable) or Soft (sand mixtures or forest ground, deformable). The right limbs signals were rotated to orientate them the same way as the left limbs', to later average their extracted features into Fore and Hind limb features. The data were segmented using a sliding window of 200 samples with a 50% overlap. For each signal from each window, time-domain features (min, max, mean, standard deviation, skewness, kurtosis, first and third quartiles), positive and negative peaks- and zero-cross-counts, and frequency-domain (spectral entropy and energy, magnitude, and phase angle of the first six Fourier transform coefficients) were extracted.

2.3. Feature selection algorithms

The FSAs used were chosen from the *Feature Selection Library v7.0.1_2020_2* (Roffo 2018). To constitute the benchmark, a panel of ten FSAs were selected among the nineteen available, based on supervised and unsupervised methods: INFFS-supervised, INFFS-non supervised, ILFS, UDFS, ECFS, LLCFS, LASSO, CFS, UFSOL and Fisher. The algorithms' abbreviations and descriptions can be found in Roffo (2018).

Table 1. Percentages of features ranked among the first 100 features by each method, counted per biomechanical origin (mean(SD) of 10 iterations).

	WALK – Body features					
	Accelerations			Rotations		
	DV	CC	ML	Yaw	Roll	Pitch
ILFS	2.8(1.1)	0.7(1.1)	0.1(0.3)	0.1(0.3)	0.3(0.5)	0.1(0.3)
LLCFS	8(0.0)	8(0.0)	8(0.0)	8(0.0)	8(0.0)	8(0.0)
	WALK – Limb features					
	Accelerations			Rotations		
	PD	S	ML	PR	AA	IE
ILFS	23.8(0.6)	14.1(1.1)	15.8(1.1)	3.4(1.0)	5(1.2)	8.2(0.9)
LLCFS	3.9(0.9)	6.6(1.0)	3(0.8)	6(0.8)	2.4(0.8)	4(0.5)
	TROT – Body features					
	Accelerations			Rotations		
	DV	CC	ML	Yaw	Roll	Pitch
ILFS	8.3(1.1)	0.5(1.1)	0(0.3)	0.1(0.3)	2.8(0.5)	0.7(0.3)
LLCFS	7.9(0.0)	7.9(0.0)	8(0.0)	8(0.3)	8(0.5)	7.9(0.3)
	TROT – Limb features					
	Accelerations			Rotations		
	PD	S	ML	PR	AA	IE
ILFS	12.1(1.2)	12(0.6)	16.5(1.1)	1.1(1.0)	5.2(1.2)	4.3(0.9)
LLCFS	2.5(0.9)	3.7(1.0)	8(0.8)	7.5(0.8)	3.7(0.8)	3.4(0.5)

DV: dorsoventral; **CC**: craniocaudal; **ML**: mediolateral; **PD**: proximodistal; **S**: sagittal; **PR**: protraction-retraction; **AA**: abduction-adduction; **IE**: internal-external.

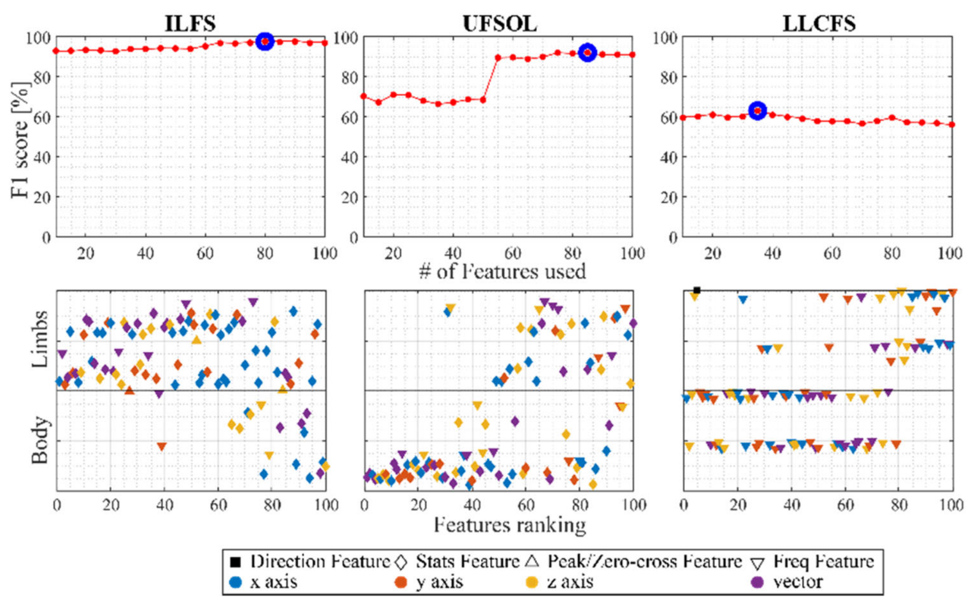

Figure 2. Visualisation of the F1-scores obtained for each FSA (column) for the trot datasets (top row), and which features nodes, axes and types were used (bottom row). Blue circles show the best results.

2.4. Dataset construction, machine learning training and evaluation

The segmented windows of 76 horses were used for training, and those of 20 horses for testing. Training and testing features were scaled, based on the training dataset. Each gait dataset was passed into each FSA to obtain a features ranking vector. The first 10 features were used to train a linear Support Vector Machine (SVM) model to classify Hard and Soft data. The model results metrics were then calculated. The same process was repeated 20 times after incrementing the number of features used to train the SVM by 5, reaching 100 training features. Next, for each FSA and gait dataset, the first 100 features ranked were categorised per node location (Body: withers, sacrum; Limbs: front, hind), sensor (accelerometer, gyroscope), biomechanical orientation and feature type (time-domain, peak/zero-cross counting, frequency-domain, gait direction). This process was repeated 10 times with different horses distribution in training and testing datasets. The features' distributions were then averaged (Table 1).

3. Results and discussion

For brevity, only the trot results of the best (ILFS) and worst (LLCFS) FSAs are presented, as well as of the UFSOL for its interesting rapid increase in F1-score around 50 features (Figure 2). FSA's ranking including Limbs features (ILFS, UFSOL after the 50st features) provided good classifications at both gaits (F1-scores >80%), whereas FSA's ranking including mainly Body features (LLCFS) provided worse classifications (F1-score <65%). With the ILFS ranking, at walk, the proximo-distal accelerations features were the most used for good classification, while at trot, medio-lateral acceleration features were prevalent (Table 1). The gyroscope features were sparsely used. At trot, horses limbs act like a spring-mass model, damping the impacts during locomotion (Wilson et al. 2001). It was thus expected that the differences in signals between Hard and Soft surfaces would be found in the limb features. However, human literature has shown that it is possible to classify outdoor terrains encountered by runners with only a lower-back 3D accelerometer (F1-scores >88%) (Dixon et al. 2019). Our findings show that using only one body node would give insufficient results in horses, especially at walk.

4. Conclusions

Comparing the features ranking of different FSA and visualising the provenance of the selected features against the trained models' evaluation metrics is of high value to better understand the relation between the input and the outputs in a classification task. Moreover, our method can be applied to different biomechanical datasets. Our methods show that new biomechanical insights can be gained by investigating features chosen by classification models and, therefore, learning from previously considered black-box methods.

Disclosure statement

No potential conflict of interest was reported by the authors.

Acknowledgements

The authors thank the horses, owners, the people involved in data collection, the UU and UT for funding JIMP's research project. Photo: courtesy of Ineke H. Smit (UU).

References

Bosch S, Serra Bragança F, Marin-Perianu M, Marin-Perianu R, Van der Zwaag BJ, Voskamp J, Back W, Van Weeren R, and Havinga P et al. 2018. EquiMoves: a wireless networked inertial measurement system for objective examination of horse gait. Sensors. 18(3):850.

Dixon PC, Schütte KH, Vanwanseele B, Jacobs JV, Dennerlein JT, Schiffman JM, Fournier PA, and Hu B et al. 2019. Machine learning algorithms can classify outdoor terrain types during running using accelerometry data. Gait Posture. 74:176–81.

Roffo G. 2018. Feature Selection Library (MATLAB Toolbox). [Accessed August 16, 2021]. https://arxiv.org/abs/1607.01327

Serra Bragança FM, Broomé S, Rhodin M, Björnsdóttir S, Gunnarsson V, Voskamp JP, Persson-Sjodin E, Back W, Lindgren G, Novoa-Bravo M, and Gmel AI et al. 2020. Improving gait classification in horses by using inertial measurement unit (IMU) generated data and machine learning. Sci Report. 10:17785.

Wilson AM, McGuigan MP, Su A, and Van den Bogert AJ. 2001. Horses damp the spring in their step. Nature. 414(6866):895–99.

KEYWORDS Horse; biomechanics; machine learning; feature selection algorithm; IMU

j.i.m.parmentier@uu.nl