# Non-Invasive Lactate Estimation Using Wearable Sensors for Remote Fatigue Assessment in Horses

Hamed Darbandi
*Pervasive Systems Group*
*University of Twente*
Enschede, The Netherlands
h.darbandi@utwente.nl

Carolien Munsters
*Equine Integration B.V., Hoogeloon*
*Faculty of Veterinary Medicine, Utrecht Uni.*
Utrecht, The Netherlands
carolien@munsters.nl

Paul Havinga
*Pervasive Systems Group*
*University of Twente*
Enschede, The Netherlands
p.j.m.havinga@utwente.nl

2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) | 979-8-3503-0436-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/PerComWorkshops59983.2024.10503247

*Abstract*— **Exercise-induced fatigue is a complex phenomenon that can significantly impact the health and welfare of horses. Traditional methods for assessing fatigue in horses, such as plasma lactate accumulation (LA) measurement, can be invasive and require the presence of a veterinarian on-site. In this paper, we propose the use of body-mounted inertial measurement units (IMUs) and a heart rate (HR) monitor as a non-invasive and veterinarian independent approach for assessing fatigue by estimating LA in horses during exercise. LA estimation models were trained using signal-based features and kinematic parameters extracted from IMUs. As an outcome, the accuracy of the best performing model based on two IMUs and HR was 0.11 mmol/L and 4.89% (root mean square error and mean absolute percentage error). This approach demonstrates the potential for remote health monitoring in animals, which can be particularly valuable for those in remote locations or with limited access to specialized veterinary care.**

***Keywords*— *Fatigue, Horses, Inertial measurement unit, Machine learning***

## I. Introduction

Fatigue in horses refers to a state of reduced physical and physiological capacity as a result of prolonged or intense exercise. This state can significantly impact the overall welfare of horses, potentially leading to adverse effects on multiple body systems, including muscular, cardiovascular, and respiratory functions [1]. Incorporating fatigue management into equestrian sports is crucial for the well-being of horses. By understanding the physical demands placed on these animals and actively monitoring and addressing fatigue, trainers and riders can contribute to the longevity, health, and success of equine athletes while upholding ethical standards and public opinion [2].

Human athletes often express their fatigue level based on a rate of perceived exertion (RPE), which can be used as reliable feedback for managing the exercise intensity and assessing fatigue [3]. In contrast, horses cannot verbally express their fatigue state. Since fatigue is a multidimensional phenomenon, there is no single definition that is universally accepted across equine studies. In many studies, fatigue has been defined when a horse is unable to maintain its pace on a treadmill despite verbal encouragement [4]. This definition is practical in a controlled setting, but it is not feasible for on-field exercise or competition as other factors like rider influence come into play.

Traditionally, blood lactate accumulation (LA) measurement during exercise has been used for fatigue evaluation in human athletes [5]. Heart rate (HR), while informative, can be influenced by various factors. LA serves as a more direct indicator of metabolic stress and anaerobic activity, offering insights into muscle fatigue that HR alone may not always capture effectively. However, the measurement procedure for horses poses significant challenges, as it requires the presence of a veterinarian for the multiple invasive blood samples extraction from the jugular vein. Besides causing discomfort and the dependance on veterinary assistance, repeated blood samplings during exercise disrupt an optimized training session [6].

Therefore, there is a need for a non-invasive method that can replace on-site LA measurement, ensuring a thorough evaluation of fatigue. We propose the use of inertial measurement unit (IMU) combined with HR monitors as a solution for accurately estimating LA in horses. HR monitors are used during exercise to track cardiovascular responses, optimizing workout intensity, and evaluating recovery [7]. IMU sensors offer the ability to capture the movement with high precision. They have been designed for continuous measurement as opposed to the discrete measurement of LA. They are small, non-invasive, and easily mountable on the body. By analyzing their output signals, biomechanical parameters can be calculated and monitored. By monitoring key biomechanical parameters, we will be able to measure horse's fatigue levels during exercise. This approach has the potential to enhance the practicality of fatigue assessment in horses during exercise, particularly in situations where access to veterinary care is limited.

## II. Background and Prior Work

The utilization of machine learning (ML) techniques is becoming more prominent in equine studies. To evaluate different aspects of horse health and fitness, studies have used various features including time- and frequency-based features to develop ML models [8,9]. Overall, both types of features convey significant information and can help in the development of models.

In contrast to the numerous exercise fatigue estimation studies on human, there is only one study on horses. In that single study, Darbandi et al. [6] used ML algorithms to classify whether a horse is fatigued with 95 percent accuracy during walk. The most significant features from that study were stride duration, stance duration, stride length, speed, and limb range of motion. The first two increased, while the latter three decreased due to fatigue. However, their model was not able to quantify consequent fatigue changes during exercise.

Many studies on human fatigue estimation used RPE as the fatigue indicator. Aguirre et al. [10] used the data from ambulatory sensors to classify the fatigue from sit-to-stand exercise, which presented a 82.5% accuracy. Jiang et al. [11] trained a model for fatigue prediction and quantification based on the data from IMU using random forest (RF) and Convolutional neural networks (CNNs). Their model showed a high correlation for a subject-dependent regression model for continuous fatigue detection. Nevertheless, the absence of a metric comparable to RPE for horses, making it unfeasible to consider it for the assessment of exercise intensity and fatigue.

In this study, we compared the performance of different ML algorithms for modeling fatigue using motion and HR data. To overcome the absence of RPE, we considered LA as the fatigue indicator. Furthermore, we examined the influence of the IMU placement on the horse's body on the performance of the estimation model.

## III. Methods

### A. Dataset Preparation and Feature Extraction

The Data were collected from twenty-one ridden eventing horses (11.9 ± 2.9 years old). They were equipped with two ProMove-mini IMUs [12] on the sacrum and the limb cannon bone (Fig. 1). Each IMU contained a tri-axial accelerometer and a tri-axial gyroscope and was set to collect data at a sampling rate of 200 Hz, an acceleration range of ±8g, and an angular velocity of ±2000 deg/s. These two sensors generated two three-dimensional acceleration and angular velocity signals.

IMU locations and orientations on the body were demonstrated in Fig. 1. The three axes of rotation for the sacrum IMU were x, y, and z, which were defined in the order as longitudinal axis (or pelvis roll angle), mediolateral axis (or pelvis pitch angle), and vertical axis (or pelvis yaw angle). For limb, x-axis was aligned to the cannon bone and external hoof wall (vertical axis or internal/external rotation angle), while y- and z-axis were set as longitudinal axis (or abduction/adduction angle) and mediolateral axis (or Protraction/Retraction angle).

The horses were measured while performing a standard exercise test (SET). SET is a type of exercise test that is used to assess the fitness level of horses. The SET involves a step-wise approach gradually increasing the intensity of exercise according to a protocol until the horse reaches a predetermined LA threshold. The anaerobic threshold, also known as the LA threshold (often at 4 mmol/L), signifies the point where body's capacity to clear LA from the bloodstream is surpassed. As this threshold is crossed, there is a swift escalation in LA levels, serving as an indicator that the horse is nearing fatigue [7].

The SET protocol in this study involved ridden walking, trotting, and cantering. It was designated with a LA threshold of around 4 mmol/L, and consisted of the following steps: 8-10 minutes of walk and 8-10 minutes of trot for warming up, four 2-3 minutes of incremental steps of cantering with speeds of 5.6, 6.8, 8.0, and 9.2 m/s, and finally, 5-6 minutes of cool-down walk. Within 15 seconds after each incremental steps and the


Fig. 1. IMUs locations and orientations on horse body.

cool-down, a blood sample was taken from the jugular vein and a portable hand-held measurement device (Lactate Pro2, Arkray Inc., Kyoto, Japan) was used to determine the plasma LA. If the plasma LA exceeded 4 mmol/L, The subsequent incremental steps were omitted, and the SET was terminated with cool-down [13]. The LA values from the SET are demonstrated in Fig. 2.

In addition to LA, HR was recorded with a sampling frequency of 1 Hz using an Polar HR monitor (Polar Electro, Finland) equipped on the horses throughout the SET. All the subjects were examined for lameness and injuries before the study by a veterinarian. The ones that presented lameness or injuries during the examinations were excluded from this study.

The collected data from IMUs, HR monitor, and plasma LA measurements were time-synchronized. All the six signals were used from the sacrum IMU, while only the three angular velocity signals were used from the limb IMU. The acceleration signals from limb IMU were excluded due to the high exercise intensity, which caused the acceleration values to surpass the ±8g threshold of the IMU's initial settings.

The IMUs signals were smoothed using Savitzky-Golay polynomial filter. Since there is at least one cantering stride within one second, the filtered signals were windowed into windows wider than one second to cover at least one full stride. To study the impact of window size on the model, we chose two, four, and eight-second windows (400, 800, and 1600 samples). Consecutive windows were extracted with a 25 percent overlap.

Feature extraction was implemented on the windows for the purpose of discovering the variables that significantly change the most by fatigue. The extracted features consisted of eleven signal-based features (per signal) and fifteen kinematics parameters (in total), as presented in Table I.

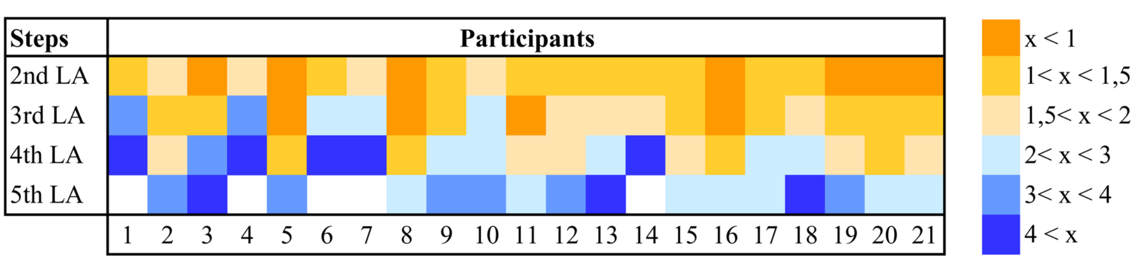


Fig. 2. LA values of participants. The 5th LAs of participants 1, 4, 6, 7, and 14 are empty (white color) due to reaching 4 mmol/L in 4th LA.

The signal-based features represent different signal aspects. For instance, standard deviation was the most effective variability metric selected for sport horses fatigue classification [6]. Additionally, an important feature in distinguishing the rider's impact on horse motion was the magnitude of the first fast Fourier transform (FFT) coefficient of sacrum acceleration signal [8]. The frequency features were extracted from the FFT using the von Hann function to minimize spectral leakage [6,8].

Fifteen kinematic parameters were also calculated (Table I). These parameters have been extensively discussed in equine gait literature and are vital for evaluating health and fitness [6,8,9]. Speed and stride, stance, and swing duration were selected, which have been used in fatigue literature, especially where the effect of exercise was the focus of the study [6]. To accurately estimate the gait events, a gait event detection algorithm was used [15]. Speed was estimated using a horse speed model [8], which takes signals from sacrum or limb IMUs as input and generates the corresponding speed value.

Due to the SET protocol, stride temporal parameters and speed inevitably increase throughout the SET, which might not reflect the true value of them. To have a normalized parameter that is not affected by the intensity and increment of the parameters, we define Normalized Stride-Efficiency Factor (NSEF), which is calculated as stride duration divided by speed.

Angular range of motion (AROM) was calculated by integrating the angular velocity signals per stride, resulting in an angle signal [6]. AROM for each stride was determined by calculating the difference between the maximum and minimum values of the angle signal. The assessment of AROM involved considering the limb orientation around three axes with the cannon bone as the limb and the carpal joint as the reference point, as elaborated in Section A of Methods. Additionally, AROMs of the pelvis around three axes were measured using the IMU's center as the rotation reference point, as shown in Fig. 1.

Linear range of motion (LROM) was assessed through the analysis of displacement signals obtained from the IMUs across all strides. We calculated the mediolateral and vertical LROM for both limb and sacrum by determining the difference between the maximum and minimum displacement values within each stride. The displacements were calculated by employing a cyclical integration process on the acceleration signals, following the methodology described in [12].

MaxDiff and MinDiff are defined as the difference between the two maximum (MaxDiff) and two minimum peaks of (MinDiff) sacrum vertical displacement within a single stride. These two parameters have been used as one of major metrics for identifying lameness in horses [12].

The IMU signals were rigidly windowed, however, most of the kinematic parameters are based on a full stride. To address this, we utilized the stride event detection algorithm [15] to extract the maximum possible number of strides from each window, and subsequently calculated the kinematic parameters from the stride(s) within the window. In cases where more than one stride occurred within a window, we considered the average value of the kinematic parameters.

*B. LA Calculation*

As mentioned, only six LA data were collected per horse during SET. However, more LA data per horse was required for implementing an LA estimation model. Hence, we converted the discrete LA measurements into a continuous format. In the literature, a relationship has been established between LA and HR [16, 17],

$$y = be^{cx} \tag{1}$$

This formula is participant-dependent, which derives from the ΔHR-LA response curve, where x is fractional elevation of HR ($\Delta HR$) and y is LA. $\Delta HR$ comes from:

$$\Delta HR = \frac{HR_{exercise} - HR_{rest}}{HR_{maximal} - HR_{rest}} \tag{2}$$

To fit the $\Delta HR$-LA curve, a SET with increasing intensity (increasing LA and $\Delta HR$) is required. Therefore, we excluded the first (at rest) and the sixth (after recovery) LA samples from the $\Delta HR$-LA plot, which left a maximum of four LA samples per participant (Fig. 2). Then, exponential curve was fitted to the four (or fewer) $\Delta HR$ and LA datapoints for each horse. An illustrative example of a fitted line in $\Delta HR$-LA plot for one of the participants is demonstrated in Fig. 3. Using the fitted curve, the individual-dependent factors, b and c, were extracted [16]. With the formula and continuous $\Delta HR$ data from the SET, we

TABLE I. SIGNAL-BASED AND KINEMATIC PARAMETERS

| Parameter description | Extracted from | | |
|---|---|---|---|
| | Sacrum | Limb | Both |
| ***Signal-based features*** | | | |
| Maximum, Minimum, Mean, Median | 24 | 12 | 36 |
| Standard deviation, First and third quartiles | 18 | 9 | 27 |
| Detrended Fluctuation Analysis | 6 | 3 | 9 |
| Spectral energy [a] [8] | 6 | 3 | 9 |
| Magnitude of the first three coefficients of FFT | 6 | 3 | 9 |
| Stride regularity based on sacrum IMU [14] | 1 | - | 1 |
| Stride regularity based on limb IMU | - | 1 | - |
| ***Kinematic parameters*** | | | |
| Stride duration, stance, and swing duration, Speed | 4 | 4 | 8 |
| Normalized Stride-Efficiency Factor (NSEF) | 1 | 1 | 2 |
| Limb AROM | - | 3 | 3 |
| Sacrum AROM and LROM, MaxDiff, MinDiff | 7 | - | 7 |
| **Total** | 73 | 39 | 112 |

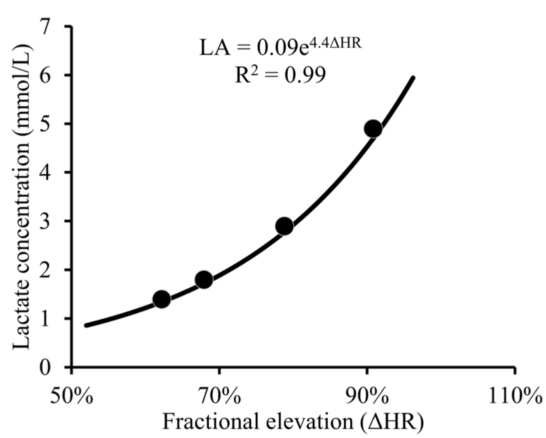


Fig. 3. $LA_g$ plotted against the fractional elevation of HR of a participant. Exponential line provides calculation of the weighting factor (y).

calculated an LA, which we assigned it as the generated LA ($LA_g$). $LA_g$ was used for the estimation model development.

A resting heart rate of 30 and a maximum heart rate of 220 beats per minute were employed for ΔHR assessment [16] as it was not possible to obtain the maximum heart rate during this submaximal exercise test or accurately measure the resting heart rate in these high-performance sport horses in the field.

### C. *Model Training, Optimization, and Evaluation*

The dataset was randomly divided into fourteen two-member groups. In each iteration, one group served as the testing set while the remaining horses formed the training set. This approach enabled robust model training with combination of various horses. After each split, the training set was Z-normalized. Then, the testing set was Z-normalized using the parameters from the training set Z-normalization to remove the bias from the testing set.

Deep learning (DL) techniques were applied on the training sets as input and $LA_g$ as output. The training sets were features from 1- HR only, 2- HR and both IMUs (IMU+HR), 3- both IMUs, 4- Sacrum IMU only, and 5- limb IMU only. This has been done to assess the impact of HR as a physiological feature on the model accuracy and to identify the compromised accuracy on the single-IMU models due to less features. All the models were fourteen-fold trained.

CNNs and recurrent neural networks (RNNs) with a long short-term memory (LSTM) architecture were selected for their proven effectiveness in forecasting multiple time series across various settings. CNNs excel at capturing temporal patterns effectively assuming the same patterns are relevant throughout the data. They also exhibit computational efficiency compared to RNNs. On the other hand, RNNs were chosen for their capacity to forecast multiple time series.

We compared DL models with the following baseline models: random forest (RF), neural networks (NN), and decision tree (DT). Notably, for the HR-only dataset, we omitted CNN and LSTM models due to the excessive complexity of training a sophisticated model with a single input feature.

The accuracy and relative error of the estimation results were evaluated using root mean squared error (RMSE, Eq. (3)) and mean absolute percentage error (MAPE, (Eq. (4)). To achieve better model performance, we tuned several important hyperparameters via a grid search before training the models. The activation function in all CNN, LSTM, and NN models was set to ReLU (Rectified Linear Unit), while the batch size was considered 16. The details of the hyperparameter search space for the models are demonstrated in Table II.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}} \quad (3)$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{Predicted_i - Actual_i}{Actual_i}\right| \quad (4)$$

Finally, to identify the most impactful features on fatigue estimation, we conducted a feature search to identify their relevance. Hence, the top ten features from Neighborhood Component Analysis (NCA) were selected for further study.

TABLE II. HYPERPARAMETER SEARCH SPACE

| Process/Model | # | Hyperparameter | Range of values |
|---|---|---|---|
| Data preparation | 1 | Window size | 2, 4, 8 seconds |
| CNN | 1 | Number of units, Number of filters | 16, 32, 64, 128 |
| | 2 | Kernel size, Pool size | {3, 5, 7}, {2, 3} |
| | 3 | Number of units in dense layer | 16, 32, 64, 128 |
| LSTM | 1 | Number of units | 25, 50, 75, 100,150 |
| NN | 1 | Number of units in hidden layer(s) | ($N^a$/4), (N/2), (N), (N/2,N/4), (N,N/2) |
| | 2 | Alpha (L2 regularization) | 0.01, 0.001, 0.0001 |
| RF | 1 | Maximum depth of tree | None, 5, 10, 15 |
| | 2 | Maximum features for best split | $N^2$, $Log_2N$ |
| | 3 | Minimum samples for a leaf node | 2, 5, 10 |
| | 4 | Number of estimators | 50, 100, 200, 500 |
| DT | 1 | Maximum depth of tree | None, 5, 10, 15 |
| | 2 | Maximum features for best split | $N^2$, $Log_2N$ |
| | 3 | Minimum samples for a leaf node | 2, 5, 10 |

[a] N is the number of features

## IV. RESULTS

All the processes and calculations were performed using Python 3.7.0 with the TensorFlow 2.4.0 and PyTorch 2.0.1 modules. The total amount of data was 16,500 seconds (or 3,300,000 samples). The estimation accuracy of the trained models and the optimized hyperparameters are presented in Table III, where the hyperparameters were labeled with "#" as their column headers, corresponding to the "#" column in Table II. Moreover, the estimation results of the best performing models between IMU+HR and HR training sets are compared in a scatter plot (Fig. 4). We showcased the most significant features based on both IMUs dataset in Table IV. Moreover, using ΔHR and the coefficients (b, c) derived from the fitting, the $LA_g$ of each participant was determined independently.

DL models performed better than the baseline, where the best and worst performing models were based on CNN and NN, respectively. The range of performance metrics were 0.11-0.98 mmol/L for accuracy and 4.51 to 46.11% for relative error. Setting the window size to four seconds resulted in improved performance for all the models.

The CNN and LSTM models based on IMU+HR training set yielded the best accuracy (0.11 mmol/L) and lowest relative error (4.51%), as shown in Table III. Between models based on single IMU, the CNN model based on the sacrum IMU presented higher accuracy and lower relative error (0.25 mmol/L, 9.6%) than the limb IMU (0.32 mmol/L, 11.79%). Furthermore, the model accuracy increased whenever HR was added to the IMU features (Best: CNN, 0.11 mmol/L, 4.89%). In contrast, the accuracy decreased when HR was considered as the only input feature (Best: RF, 0.39 mmol/L, 13.73%).

The ten most relevant features for $LA_g$ estimation are presented in Table IV. Five and three out of the ten features were relied on the sacrum and limb IMU, whereas two features were independent of the IMU (i.e. extractable from sacrum or limb IMU). Moreover, three features were kinematics parameters, while the remaining were signal-based features.

TABLE III. BEST SET OF HYPERPARAMETERS AND PERFORMANCE METRICS OF THE TRAINED MODELS

| Model | Input | Hyperparameteres | | | | RMSE (mmol/L) | MAPE (%) |
|---|---|---|---|---|---|---|---|
| | | *#1* | *#2* | *#3* | *#4* | | |
| CNN | IMU+HR | 32,128 | {7},{2} | 128 | - | 0.11 | 4.89 |
| | Both IMUs | 32,128 | {7},{2} | 128 | - | 0.22 | 8.72 |
| | Sacrum IMU | 128,128 | {5},{3} | 128 | - | 0.25 | 9.60 |
| | Limb IMU | 64,64 | {7},{2} | 128 | - | 0.32 | 11.79 |
| LSTM | IMU+HR | 75 | - | - | - | 0.14 | 4.51 |
| | Both IMUs | 75 | - | - | - | 0.26 | 9.70 |
| | Sacrum IMU | 100 | - | - | - | 0.30 | 11.95 |
| | Limb IMU | 50 | - | - | - | 0.38 | 13.62 |
| NN | HR | - | - | - | - | 0.79 | 28.54 |
| | IMU+HR | (N) | 0.0001 | - | - | 0.74 | 35.14 |
| | Both IMUs | (N) | 0.001 | - | - | 0.80 | 38.11 |
| | Sacrum IMU | (N) | 0.001 | - | - | 0.86 | 39.73 |
| | Limb IMU | (N) | 0.0001 | - | - | 0.98 | 46.11 |
| RF | HR | - | - | - | - | 0.38 | 13.73 |
| | IMU+HR | 10 | $N^2$ | 2 | 50 | 0.24 | 9.85 |
| | Both IMUs | None | $N^2$ | 2 | 500 | 0.29 | 10.38 |
| | Sacrum IMU | None | $N^2$ | 2 | 200 | 0.33 | 11.65 |
| | Limb IMU | None | $N^2$ | 2 | 100 | 0.39 | 13.79 |
| DT | HR | - | - | - | - | 0.40 | 13.78 |
| | IMU+HR | 10 | $N^2$ | 5 | - | 0.39 | 13.81 |
| | Both IMUs | 10 | $N^2$ | 10 | - | 0.46 | 19.73 |
| | Sacrum IMU | None | $N^2$ | 8 | - | 0.52 | 22.55 |
| | Limb IMU | 10 | $N^2$ | 2 | - | 0.54 | 21.76 |

## V. DISCUSSION

This paper proposed a data-driven method to evaluate the exercise-induced fatigue of sport horses by estimating the LA non-invasively using wearable IMU and HR data. It is important to emphasize that the proposed model is participant-independent, meaning it can generalize well to new horses.

Estimating the LA during exercise provides riders and trainers with capability to monitor the fatigue level of the horse, helping them make informed decisions about whether to continue or halt the exercise to prevent potential injuries or overtraining Furthermore, during a SET, they can seamlessly record important performance measures, such as speed and HR at the point when LA reaches a specific threshold [7]. The most accurate model in this study yielded a RMSE of 0.11 mmol/L, which was biologically significant and precise for the LA measurements around 4 mmol/L (the range of LA in this study).

According to Table III, the models based on sacrum IMU demonstrated higher accuracy and lower relative error compared to the models based on the limb IMU, independent of ML technique. This suggests that sacrum IMU captured and conveyed more fatigue-related information than limb IMU. It's worth noting that the difference might also arise from the disparity in the number of input features between limb IMU and sacrum IMU models—approximately twice as many in the latter case (as explained in the Methods, limb IMU acceleration signals were excluded due to exceeding the specified threshold in IMU initial settings).

As anticipated, the performance of the models improved as the number of input features increased, following the pattern: limb IMU < sacrum IMU < both IMUs < IMU + HR. Specifically,

TABLE IV. THE TEN HIGHEST RANKING FEATURES FROM NCA

| # | Feature description | IMU | Signal | Axis/Angle |
|---|---|---|---|---|
| 1 | Spectral energy | Sacrum | Ang. Velocity | Roll angle |
| 2 | Spectral energy | Limb | Ang. Velocity | Protraction/Retraction |
| 3 | NSEF | - | - | Longitudinal axis |
| 4 | AROM | Limb | - | Abduction/Adduction |
| 5 | Mag. FFT 1st 3 coefficients | Sacrum | Ang. Velocity | Yaw angle |
| 6 | Spectral energy | Sacrum | Acceleration | Vertical axis |
| 7 | AROM | Limb | - | Protraction/Retraction |
| 8 | Stride regularity | Sacrum | Acceleration | Vertical axis |
| 9 | Standard deviation | Sacrum | Acceleration | Longitudinal axis |
| 10 | Maximum | Sacrum | Acceleration | Vertical axis |

the addition of HR to the training set resulted in an improvement in accuracy by 0.11 mmol/L and a reduction in relative error by 3.83% (CNN model). However, there was a noteworthy exception: the models trained solely on the HR outperformed IMU-based training sets within DT and NN. The reason might be attributed to the incorporation of HR data into the output. Despite HR not playing a direct role in $LA_g$ calculation and considering that LA was fitted to an exponential curve in relation to HR, there could be resemblances in their patterns. Furthermore, employing only two IMU devices on the horse body for $LA_g$ estimation results in better accuracy (RF model: 0.24 mmol/L) compared to using only a HR monitor (RF model: 0.38 mmol/L).

CNN, LSTM, and RF models yielded superior results compared to DT and NN. The disparity in accuracy among models based on IMU+HR training set was more pronounced in the case of CNN and LSTM (0.11 and 0.12), as opposed to RF where it was only 0.05 mmol/L. However, the accuracy of

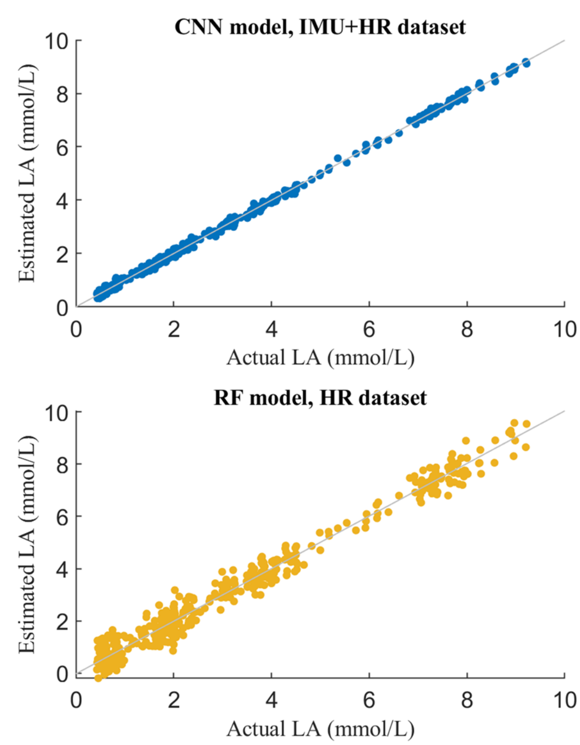


Fig. 4. Examples of estimated LA over a participant SET from (above) CNN model, IMU+HR training dataset and (below) RF model, HR training set

models that relied solely on IMU data demonstrated similar performance across these three algorithms. This observation may stem from the increased complexity when handling two distinct data types, motion and physiological data, by LSTM and CNN. Conversely, RF exhibits lower computational complexity compared to DL models. This aspect emphasizes the utility of RF over CNN and LSTM in scenarios with constrained resources and the absence of an HR monitor.

Spectral energy has been selected three times as one of the top ten relevant features (Table IV). Changes in the spectral energy distribution can indicate the presence of abnormal patterns in the signal. The spectral energy of pelvis roll angular velocity can be used as an indicator for pelvis roll angle, which has also been utilized as a metric for lameness [12]. A similar approach could also be utilized for the second ranked feature, where the changes in limb protraction/retraction angle, has been selected as one of the most indicative parameters in a recent equine fatigue study [6]. Therefore, the selection of these parameters as stated in the literature as performance deterioration factors might also validate their relevance to fatigue.

The inclusion of limb AROMs for protraction/retraction and abduction/adduction among the top features substantiates the findings of [6], which indicated significant fatigue-related associations for both parameters. Furthermore, standard deviation was selected as the ninth top relevant feature, meeting our expectations as an important variability parameter for IMU signals in fatigue assessment [6].

The stride regularity extracted from the sacrum acceleration signal was ranked higher than that from the limb's angular velocity signal. This observation supports the validity of the stride regularity calculation method documented in the literature [14], where the sacrum acceleration signal was employed.

## VI. Conclusion

Three significant advancements were realized for the first time in this research's scope. First, the estimation of LA was successfully executed using DL and ML models. Second, the creation of a participant-independent LA estimation model marked a noteworthy achievement. Last, an analysis of the utilization of kinematic parameters in the estimation of a physiological parameter, LA, was conducted. This study underscores the potential of leveraging data from wearables on animals, facilitating remote care and reducing the dependence on invasive fatigue measurement. The integration of a HR monitor enhances the performance of the models. The application of IMUs not only benefits riders and trainers by remotely monitoring fatigue but also operates independently of the presence of a veterinarian, contributing to the overall well-being of horses.

## Acknowledgment

The authors wish to thank Jeanne Parmentier and Esther Siegers from Utrecht University, and all the people that helped with the data collection. EFRO OP-Oost and Equine Integration B.V. partly funded this study.

## References

1. J. L. Taylor, S. C. Gandevia, “A comparison of central aspects of fatigue in submaximal and maximal voluntary contractions,” J Appl Physio, vol. 104, no. 2, pp. 542-550, 2008, doi:10.1152/japplphysiol.01053.2007
2. F. Arfuso, et al., "Peripheral Modulators of the Central Fatigue Development and Their Relationship with Athletic Performance in Jumper Horses," Animals (Basel), vol. 11, no. 3, pp. 743, Mar 2021, doi:10.3390/ani11030743.
3. R. Carrie, “Rating of perceived exertion (RPE),” J Physiother., vol. 58, no. 1, p. 62, 2012. doi:10.1016/S1836-9553(12)70078-4
4. J.R. Bowers and R.F. Slocombe, "Influence of Girth Strap Tensions on Athletic Performance of Racehorses," Equine Vet J Suppl, no. 30, pp. 52-56, Jul 1999, doi: 10.1111/j.2042-3306.1999.tb05188.x
5. G. Theofilidis, G.C. Bogdanis, Y. Koutedakis, and C. Karatzaferi, "Monitoring Exercise-Induced Muscle Fatigue and Adaptations: Making Sense of Popular or Emerging Indices and Biomarkers," Sports, vol. 6, no. 4, pp. 153, 2018, doi:10.3390/sports6040153
6. H. Darbandi, C. Munsters, J. Parmentier, and P. Havinga, "Detecting Fatigue of Sport Horses with Biomechanical Gait Features Using Inertial Sensors," PLOS ONE, vol. 18, no. 4, e0284554, 2023, doi:10.1371/journal.pone.0284554
7. C.B.M. Munsters, et al., "Exercise testing in Warmblood sport horses under field conditions," Vet J, vol. 202, no. 1, pp. 11-19, 2014, doi:10.1016/j.tvjl.2014.07.019
8. H. Darbandi, et al., "Using Different Combinations of Body-Mounted IMU Sensors to Estimate Speed of Horses—A Machine Learning Approach," Sensors (Basel), vol. 21, no. 3, pp. 798, 2021, doi:10.3390/s21030798
9. H. Darbandi and P. Havinga, "A Machine Learning Approach to Analyze Rider’s Effects on Horse Gait Using On-Body Inertial Sensors," 2022 IEEE International Conference on Pervasive Computing and Communications, Pisa, Italy, pp. 248-253, doi:10.1109/PerComWorkshops53856.2022.9767228
10. A. Aguirre, et al., "Machine Learning Approach for Fatigue Estimation in Sit-to-Stand Exercise," Sensors (Basel), vol. 21, no. 15, pp. 5006, 2021, doi:10.3390/s21155006
11. Y. Jiang, et al., "A Data-Driven Approach to Predict Fatigue in Exercise Based on Motion Data from Wearable Sensors or Force Plate," Sensors (Basel), vol. 21, no. 4, pp. 1499, 2021, doi:10.3390/s21041499
12. S. Bosch, et al., "EquiMoves: A Wireless Networked Inertial Measurement System for Objective Examination of Horse Gait," Sensors (Basel), vol. 18, no. 3, pp. 850, 2018, doi:10.3390/s18030850
13. C.B.M. Munsters, et al., "Young Friesian horses show familial aggregation in fitness response to a 7-week performance test," Vet J, vol. 198, no. 1, pp. 193-199, 2013, doi:10.1016/j.tvjl.2013.07.023
14. S. Biau and E. Barrey, "Relationship between stride characteristics and scores in dressage tests," Pferdeheilkunde, vol. 20, pp. 140-144, 2004, doi:10.21836/PEM20040205
15. H. Darbandi, F.S. Bragança, B.J. van der Zwaag, and P. Havinga, "Accurate Horse Gait Event Estimation Using an Inertial Sensor Mounted on Different Body Locations," in 2022 IEEE International Conference on Smart Computing, Helsinki, Finland, pp. 329-335, 2022, doi:10.1109/SMARTCOMP55677.2022.00076
16. C.B.M. Munsters, et al., "A prospective cohort study on the acute:chronic workload ratio in relation to injuries in high level eventing horses: A comprehensive 3-year study," Prev. Vet. Med., vol. 179, p. 105010, 2020, doi:10.1016/j.prevetmed.2020.105010
17. V. Manzi, et al., "Relation between Individualized Training Impulses and Performance in Distance Runners," Med. Sci. Sports exerc., vol. 41, no. 11, pp. 2090-2096, 2009. doi:10.1249/MSS.0b013e3181a6a959