Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

BPqu: Edge-AI Breath Phase Quantifier for Real-Time Guidance in Running

Severin Bernhart

severin.bernhart@salzburgresearch.at

Salzburg Research FGmbH

Christina Halmich

Salzburg Research FGmbH

Christoph Schranz

Salzburg Research FGmbH

Rade Kutil

Paris-Lodron-University Salzburg

Research Article

Keywords: artificial neural networks, edge computing, real-time systems, sports equipment

Posted Date: September 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9996115/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

BPqu: Edge-AI Breath Phase Quantifier for Real-Time Guidance in Running

blinded blinded[1*], blinded blinded[1,2], blinded blinded[1], blinded blinded[2]

[1*]blinded, blinded, blinded, blinded, blinded, blinded, blinded.
[2]blinded, blinded, blinded, blinded, blinded, blinded, blinded.

*Corresponding author(s). E-mail(s): blinded;
Contributing authors: blinded; blinded; blinded;

Abstract

Endurance running is a popular activity due to its accessibility. However, participation is sometimes prevented by individuals experiencing respiratory problems. Monitoring breathing with body area networks can tackle these issues by tracking respiration during exercise and providing immediate, guiding feedback. Common breathing guidance systems rely on observational data from past breath cycles and consequently inherit disruptively lagging guidance interventions if breathing pattern suddenly change. Therefore, this work introduces BPqu, a deep learning algorithm for real-time breath phase estimation on edge devices that enhances new breathing guidance opportunities in running by predicting the continuous breath phase with zero latency. BPqu predicts breathing within the next running step and detects hale change events, determines inhalation and exhalation phases, and quantifies the concurrent continuous breath phase with performance metrics of 93.1%, 85.6%, and 81.5%, respectively, with zero delay. BPqu is designed to be deployed in a textile wearable breathing and stride sensor and will be applied to enhance intuitive, non-distracting breathing guidance in running, reducing respiratory distress and increasing running enjoyment, thereby promoting human well-being.

Keywords: artificial neural networks, edge computing, real-time systems, sports equipment

1 Introduction

The design of wearable systems in sports is a growing field in research driven by the rising demand for technologies that enable, improve, and augment individual sports activities, especially on physical, social, and cognitive-emotional levels. While considerable progress has been made in wearable sports system design, a research gap has been identified in sports systems concerning the cognitive-emotional aspects that influence sports participation and performance for individuals Mencarini et al. (2019) and especially for female recreational athletes Archiza et al. (2021). Thus, this work focuses on female beginner runners.

Addressing this research gap is crucial, as it highlights the need for systems that can better support individuals not just physically, but also emotionally and cognitively. Real-time biomechanical feedback systems may offer a solution by providing immediate insights in biomechanical parameters and movements to potentially alleviate the cognitive-emotionally barriers during sports activities. Currently, these real-time biomechanical feedback systems are dominant in running applications for training and performance enhancement Hribernik et al. (2022). However, despite running being a widely favored physical activity, cognitive-emotional barriers persist, especially in endurance running. For instance, some individuals avoid participating due to respiratory distress, which challenges self-motivation even though running is generally easy-to-access due to its low barrier to entry Aliverti (2016).

The enhancement of breathing efficiency is a promising approach to overcome such barriers introduced by breathing distress. Research shows that breathing efficiency can be increased by coupling breathing to the steps, i.e., locomotor-respiratory-coupling (LRC), which refers to the phase-locking between breathing and stepping rhythms in running (e.g., two steps per inhale and exhale, respectively) Daley et al. (2013). Coupled breathing helps individuals in finding their optimal breathing pattern, thereby improving performance and enhancing the enjoyment of the activity. This concept is

increasingly supported by recent advancements in understanding conscious breathing adaptations in sports science, medicine, and psychophysiology Harbour et al. (2022).

However, coupling breathing to steps can be challenging for beginners, especially when trying to find a breathing pattern that matches their fitness level. Overemphasizing breathing can even distract from the enjoyment of running and potentially ruin the experience. Thus, intuitive, non-intrusive real-time feedback based on monitored breathing pattern aids individuals in finding their coupling rhythm, such that physiological behaviors are reinforced. Hereby, wearable sensors play a crucial role in detecting breathing during exercise, providing opportunities for real-time breathing guidance, facilitating respiratory learning processes, and enhancing self-awareness Harbour et al. (2021).

This is possible, as breathing, as a homeostatic process, operates with a dynamic equilibrium and regulatory mechanisms, i.e., autonomic control processes that are activated to reduce dynamic breath-to-breath variability and reach a more stable steady-state oscillatory breathing pattern with identical consecutive breath cycles. Thus, breathing control mechanisms are predictable to some extend based on previous breathing cycles, although internal demands and external disturbances (e.g., fatigue, threats, stressors) can influence these mechanisms. Even well-regulated respiratory systems retain a degree of random variability to maintain adaptability alongside stable oscillatory patterns, which limits breathing predictability but does not eliminate it Vlemincx et al. (2013a). Together with measured stride rates, also provided by wearable sensors, the LRC ratios can be computed based on past breath and stride cycles.

Based on observational breathing and stride data, real-time feedback can be provided. Thus, a slower and more controlled breathing rate can be induced, resulting in improved psycho-physiological well-being and a more satisfying running experience Paul (2017). Two real-time breathing guiding applications based on such feedback

are *RunRhythm* Harbour et al. (2024) and *Strive* Valsted et al. (2017), which support breathing rhythm regulation by LRC Bernhart et al. (2025). Whereas *RunRhythm* provides auditory step-adherent coupled breathing guidance, *Strive* supports coupled breathing via breath- and step-adherent guidance via haptic vibrational feedback.

Due to the observational feedback based on past data in both systems (*RunRhythm* Harbour et al. (2024) and *Strive* Valsted et al. (2017)), a lag in the feedback can be observed when fast breath or stride rate changes are present. Minimizing this lag is crucial when augmented real-time breathing guidance should be provided because it requires the feedback to be delivered within the time frame of the activity to enable effective human response Postma et al. (2022), i.e., feedback must be synchronized with the performed motion for successful incorporation and feedback loop delays should constitute a fraction of human reaction times Umek and Kos (2016). In this case, it means that feedback on breathing must be performed within the correct phase of the breathing cycle, making it necessary to identify the concurrent breathing phase.

Human reaction time varies by stimulus modality, with augmented feedback (external cues) typically being faster than inherent feedback (internal perception of bodily states, such as breathing during running) Vasilyev et al. (2019). Auditory reaction times in trained individuals can range from 150–180 ms Umek and Kos (2016), and under conditions of maximal focus, can be as low as 85 ms Pain and Hibbs (2007). However, defining a precise human reaction time for breathing is challenging due to the predominantly unconscious nature of breathing perception and regulation Feigenspan (2017). While conscious breathing adaptations can tolerate a certain feedback loop delay, the acceptable limit for this delay is unexplored Bernhart et al. (2025). Accurate zero-delay breath phase quantification enable acceptable delay determination for various conscious breathing adaptations.

This work presents the implementation and evaluation of a breath phase quantifier model (BPqu) conceptualized for the application in BANs to enhance new breathing guidance opportunities. This helps individuals to find and maintain steady-state breathing in endurance running. Edge-AI solutions are applied and investigated regarding the accomplishment of on-edge and real-time requirements when DL is deployed for breathing guidance. BPqu has been designed and assessed to operate with zero-delay, showcasing its accuracy performance metrics independent of established feedback loop delays. Nevertheless, the accuracy performance metrics of BPqu can be enhanced by incorporating a feedback loop delay, provided that the accepted delay is known.

2 Related Work

Current research in breath phase prediction involves Kalman filtering, as well as machine learning and deep learning (ML/DL) approaches to predict breath phases. When using Kalman filter approaches, heuristically pre-defined parameters are required for accurate phase determination. Additionally, the accuracy of predicting continuous breath phases decreases when breathing parameters (e.g. rate, amplitude, etc.) change Ramrath et al. (2007), making it hard for the algorithm to distinguish breath signal from noise, e.g. motion artifacts. Therefore, real-time digital filters for low-latency oscillatory quantification (LLOQ) are implemented, especially in the neuroscience domain Smetanin et al. (2020), to reveal instantaneous zero-lag phase and amplitude in oscillatory signals. Thus, LLOQ algorithms could be transferred to establish zero-lag breath phase quantification.

An oscillatory phase-locked dynamic system is present when breathing in running is applied with LRC. Thus, data processing with models of coupled oscillators (DAMOCO) could be applied to extract the instantaneous breath phase if frequency

changes are slow Rosenblum et al. (2021). The application of conventional oscillatory quantification methods (e.g. DAMOCO, Kalman, adaptive, Butterworth, etc.) in breathing is imprecise for sudden breathing control mechanisms Foussier et al. (2014). However, breath control mechanisms are predictable Vlemincx et al. (2013a), allowing ML/DL algorithms to be applied for precise modeling.

ML/DL algorithms based on wearable sensors were implemented for detecting breathing patterns, i.e. normal breathing (F1-score = 92%), coughing (72%), sighing (57%), yawning (63%), and sleep apnea (87% for central sleep, 51% for obstructive sleep)), from inertial measurement units (IMU) McClure et al. (2020). Furthermore, discrete breath phase (inspiration and expiration) using lung sound sensors achieved expert agreement rates of 97% for inspiration, 87% for expiration, and 91% for both phases combined Jácome et al. (2019). Both breathing classification algorithms were implemented using convolutional neural networks and evaluated in an offline setting, without deployment or optimization for real-time application during running. Real-time breathing signal prediction algorithms were developed with adaptive neural networks based on optical motion capture data and evaluated by the normalized root mean squared error (nRMSE) Murphy and Pokhrel (2009). Negative correlations were reported between model accuracy and both prediction interval length ($0 - 500$ milliseconds (ms)) and breathing regularity Murphy and Pokhrel (2009). An average nRMSE of 0.11 for regular and 0.56 for irregular breathing with a forecast interval length of 300 ms was observed and the accuracy was improved by 5% in most cases compared to a linear filter approach Murphy and Pokhrel (2009).

Recently, Semenkov et al. Semenkov et al. (2023) proposed a temporal convolutional network (TCN) that combines long short-term memory layers with convolutional neural network layers to achieve a low-latency (< 10 ms) and accurate phase estimation ($> 90\%$ with $< 20°$) in a dynamic frequency range ($8 - 12$ Hz) of electroencephalography (EEG) signals. The TCN contains 70,000 trainable parameters, which is small

compared to other DL algorithms Semenkov et al. (2023) because ML/DL algorithms require high computing, memory and storage capabilities, wherefore, edge-computing technologies (e.g. edge-AI, tinyML, in-memory processing) are applied if complex algorithms are developed for wearable systems Jin et al. (2022).

Dynamic inference methods such as early exits, multi-exit architectures, and adaptive computation have been proposed to reduce model complexity and latency, but since these approaches yield input-dependent inference times and are less suitable for regression tasks, this work instead focuses on minimizing parameter count and floating-point operations to ensure simplicity and deployability Teerapittayanon et al. (2016).

Highly efficient computing is required because feedback loops consist of sampling, processing, and communication delays Umek and Kos (2016). Compact personal space body area networks (BAN) are applied to eliminate communication delays and reduce feedback loops, i.e., all electronic components (sensors, actuators, processors, communicators) are located on-body in one device or close to each other without communicating to cloud services Kos et al. (2019). ML/DL algorithms are cost-intensive in training, but not in their inference. Thus, pre-trained models are applied in BANs via edge-AI algorithms for filtering meaningful data in smart sensors Singh and Gill (2023). While machine learning frameworks such as PyTorch and TensorFlow Lite support the deployment of convolutional and recurrent neural networks, the constrained computing resources, memory, and power typical of Body Area Networks (BANs) pose practical challenges for their efficient real-time execution. These limitations do not prevent the use of CNNs or RNNs outright, but they necessitate significant model optimization, e.g. quantization, pruning, and architectural simplification, to meet the strict latency and energy requirements of edge-AI systems in wearable applications Singh and Gill (2023). Reduction of the number of nodes and links between nodes, sampling rates, input and output shape sizes, and converting floating- to fixed-point arithmetics are

the most important compression methods in embedded DL systems Jensen (2016). TinyML-oriented architecture designs allow sub-100KB models to perform competitively on microcontrollers Banbury et al. (2021). Neural architecture search methods yield highly efficient models, reducing parameter count and number of arithmetic operations while maintaining state-of-the-art accuracy Tan and Le (2019).

3 Methods

The intention, implementation and evaluation of BPqu is described in the following methodology subsections. Initially, the aimed hardware for the BPqu application is described (section 3.1) including the initial model architecture with heuristically determined hyperparameters (section 3.2). Consecutively, the breath phase quantification method is presented (section 3.3) and applied software tools for BPqu implementation are highlighted (section 3.4). Afterwards, the BPqu evaluation method is described including the model train and test dataset (section 3.5.1) including its preprocessing (section 3.5.2). Finally, the BPqu evaluation criteria are explained in detail (section 3.5.3)

3.1 BPqu Application

The goal of BPqu is to accurately determine the concurrent breath phase of runners in real-time with zero-delay, i.e., without any latency. Thereby, BPqu enables new sonified real-time breathing guidance opportunities by estimating and even forecasting the breath phase. To eliminate the delays caused by network communication, the offline trained BPqu is applied on edge in a microcontroller-unit (MCU; PSoC 6 (Flash: 1MB, SRAM: 288kB, CPU: max. 150MHz but expected 40MHz), Infineon Technologies) of a custom designed printed circuit board (PCB) Bernhart et al. (2024) for a tailored smart-textile wearable chest sensor for stride and breathing detection

via IMU and thoracic pressure sensing in running Bernhart et al. (2023). In addition, wired headphones are assumed for a BPqu application to neglect communication delays to actuators. The MCU applies a sampling rate of 50Hz that induces a sampling delay of 20ms. The BPqu complexity must be minimized to ensure processing delays remain below the threshold of human perception, however at the same time, the breath phase quantification accuracy must be sufficiently high to avoid disturbing runners with mismatched sonified breathing instructions. During application, all data is processed in-memory on the device and discarded immediately after the running session, ensuring no persistent data storage. Server-side data hosting (accorded to privacy standards) is optionally supported to enable application-specific post-processing. Conclusively, BPqu provides breathing guidance in running, which trades off only minimal accuracy for zero to even negative latency Ellis et al. (2013).

3.2 BPqu Architecture

The BPqu network takes two one-dimensional input signals: (a) raw breathing and (b) acceleration magnitude as norm of three raw dimensions. Both tensors are fed into a sequential of a long-short-term-memory network (LSTM) and a subsequent convolutional neural network (CNN). Therefore, the input signals are forecasted through the LSTM by capturing temporal dependencies in sequential data, such as rhythmic breathing patterns, and spatial patterns are captured in the CNN. The sequential hyper-parameters (number of LSTM layers, LSTM hidden size, number of CNN layers, CNN hidden and kernel size) are manually fine-tuned to achieve optimal performance. A learning rate of 0.001 and a batch size 128 ensures a balance between learning speed and model stability. For the LSTM and CNN, initial maximal hidden and kernel sizes of 32 were heuristically determined. Different input and output sequence lengths are evaluated. Additionally, the effect on model accuracy and complexity if channel (b) is removed as input tensor is investigated. A maximum number of 50 epochs for training

per hyper-parameter combination was selected with an automated training stop if the loss criteria converges.

3.3 BPqu Partial Estimations and Reconstruction

The stacked BPqu network aims to estimate the breath phase from breathing and acceleration measurements. The overall processing latency is compensated by predicting the breath phase with a given time range ahead. The ground truth breath phase is thereby a time-series with a specific non-continuous pattern, as depicted in Figure 1.

The universal approximation theorem states that any continuous and bounded function can be estimated by an MLP with a finite number of neurons Hornik et al. (1989). Due to the non-continuous points and asymmetry of the ground truth breath phase signal, a direct approximation of this target signal in an end-to-end ML/DL paradigm would likely lead to distortions. Consequently, a subsequent estimation of the timestamps of hale changes in this distorted approximation, i.e., the moments when runners change from inhalation to exhalation and vice-versa represented in the optima of the approximated breath phase, would be systematically biased and shifted in time. Even worse, the confidence of the model might affect the bias of estimated parameters in subsequent analysis of the approximated breath phase, such as of the detection of hale changes.

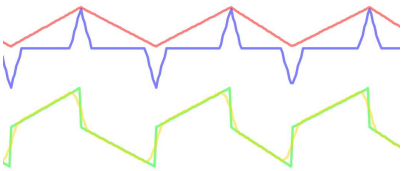


Fig. 1 Surrogate signals illustration – The two surrogate signals (base in red, event in blue) estimated by the BPqu compose the true breath phase (green) by approximation through breath phase quantification (yellow) via the reconstruction function (eq. 1).

To mitigate this systematic bias, the target breath phase $f(t)$ is decomposed into two signals. The idea is to estimate these components simultaneously by the

BPqu network and reconstruct the target breath phase from the decomposed signals subsequently. These two functions have superior properties for being approximated by an artificial neural network while preserving key characteristics, therefore they are referred to as surrogate signals. The two components, also depicted with their approximations in Figure 1, are the following:

$$f(t) = f_{\text{base surrogate}}(t) - k \cdot \int_{-\infty}^{\infty} f_{\text{event surrogate}}(t)dt \tag{1}$$

With:

- $f_{\text{base surrogate}}(t)$: A piecewise linear function between zero and one. Ones depict the beginnings of inhale phases, while zeros the beginnings of exhale phases. This base signal is illustrated in Figure 1 in red with its approximation by the BPqu network.
- $f_{\text{event surrogate}}(t) := \Sigma_{\text{inhales}}\delta'(t_{\text{inhale}}) - \Sigma_{\text{exhales}}\delta'(t_{\text{exhale}})$: A function that models the discontinuities of hale changes as sum of (directed) approximated Dirac delta functions $\delta'(t)$. While the change from exhale to inhale, i.e., the beginning of an inhale phase, is positive, the exhalation start is directed negatively. This events representing surrogate signal is illustrated in Figure 1 in blue, with its approximation by the BPqu network.

To estimate the event surrogate function $f_{\text{event surrogate}}$ by means of quantified and bounded values, each hale change event is approximated by a triangular shape $\delta'(t)$ Schranz et al. (2024). Therefore, each Dirac impulse is not approximated by a binary mask, but with a triangular symmetric kernel with a maximum of 1 at the event occurrence, a linear slope within a selected width of Δt, and zero elsewhere. This width balances the sharpness of the pulse between high precision and robustness. The factor k scales the integral of the pulses such that $k \cdot \int_{-\infty}^{\infty} \delta'(t)dt = 1$. Thus, each pulse leads to the desired sharp change of the resulting estimation with a magnitude of one.

The BPqu network estimates the two one-dimensional signals $f_{\text{base surrogate}}(t)$ and $f_{\text{event surrogate}}(t)$. While the approximation of each of these signals preserve the

inherent properties for an unbiased subsequent parameter estimation, the resulting breath phase time-series can be reconstructed easily from both components using the equation 1. The breath phase can be derived from the slope information of the estimated $f_{\text{base surrogate}}$. The hale change events can be detected from $f_{\text{event surrogate}}$ with a classical peak detection without temporal bias or influence of the model's confidence. The ground truth breath phase and its reconstruction from the two respective approximations $\hat{f}_{\text{base surrogate}}(t)$ and $\hat{f}_{\text{event surrogate}}(t)$ is depicted in Figure 1 in green.

3.4 BPqu Implementation

The MCU is programmed via ModusToolbox Technologies (2024) that can apply different inference machines: a) *ifx* inference runs Keras-H5 model formats and b) *tflite* inference runs TensorFlow-Lite model (TFLM) formats. The tflite inference can not run recurrent neural networks (RNN), i.e., LSTM and gated recurrent units (GRU). BPqu entails the LSTM, therefore, the ifx inference must be applied with a Keras-H5 model. Consequently, the BPqu implementation, training and testing was rendered in TensorFlow Developers (2024) with prequantized Integer-8bit data. Finally, the trained BPqu was quantized and compiled in imagimob AB (2024) to prepare for deployment on the MCU.

3.5 BPqu Evaluation

3.5.1 Dataset

The dataset collected for the development of the stride and breathing sensor Bernhart et al. (2023) is used for the training and testing of BPqu. Breathing and stride data were collected with a custom-made wearable device at a sampling rate of 70 Hz, with an average breath rate (BR) of 32.1 breaths per minute (bpm) within a range of $24-48$ bpm ($0.4-0.8$ Hz) and a step rate (SR) of 158.5 steps per minute (spm) within a range of $144-192$ spm ($2.4-3.2$ Hz). The runners were supposed to apply LRC, i.e.,

their BR is coupled to their SR. The LRC-ratio is determined by dividing the number of steps per breath ($\frac{SR}{BR}$). LRC-ratios of 4 and 6 were alternately instructed. Steps per inhale and exhale were equally balanced, such that a symmetric breath duty cycle can be assumed in the dataset. In total, 3h 48min of data from nineteen young (age: 20-27) female runners were recorded. The sensor signals, signal characteristics, and breathing behavior show no significant variation across gender and age groups Bloch et al. (1991), making the dataset representative of a broader demographic target population. The University Ethics board approved this study and written informed consent was obtained from all participants.

3.5.2 Preprocessing

The dataset was resampled from 70 Hz to 50 Hz with linear interpolation to optimize model training for the aimed MCU. The sample size that was used for model training and testing amounted between 57006 and 57692 samples, depending on the different tensor lengths in forecast length evaluations. To ensure the dataset represents approximately all possible running states and covers the five gait phases Novacheck (1998) in every collected stride cycle, the number of samples was set to five times the number of stride cycles. The tensor begin indices were randomly selected to avoid redundancy. The experiments for hyper-parameter optimization (hyper-opt) were rendered with a participant-wise 85 to 15 % train and test split with sample randomization per participant.

After the hyper-opt, leave-one-subject-out cross validation (LOSO-CV) is applied on the resulting BPqu model with the optimal complexity-accuracy trade-off to continue the model evaluation with breathing regularity levels in different casting conditions. Therefore, continuous predictions of surrogate signals were generated that represent the expected real-time model output at the inference time-point including the processing delay. Thus, the BPqu was applied for each sample in the dataset to reconstruct the continuous signals, resulting in 617,424 model executions per

participant in the cross validation. The continuous casting signals were calculated with respect to the delay required for processing in the number of sample indices ($\#idcs_p = \lceil \frac{t_p}{t_s} \rceil$) depending on the sampling delay (t_s) and processing delay (t_p) in ms. Figure 2 presents the casting signal reconstruction method based on the following calculations (eqs. (2), (3), (4)) where x_i are the BPqu estimated tensor values and w_i the weights:

- Backcasting: History breath phase estimation by calculating a backcast mean using a weighted sum of historical values at respective index estimations of the past h_{len} of windows:

$$backcast_{mean} = \frac{\sum_{i=1}^{h_{\text{len}}} w_i \cdot x_i}{\sum_{i=1}^{h_{\text{len}}} w_i} \tag{2}$$

$$\text{with} \quad w_i = \begin{cases} i & \text{if } i < \frac{h_{\text{len}}}{2} \\ \frac{h_{\text{len}}}{2} & \text{if } i = \frac{h_{\text{len}}}{2} \\ h_{\text{len}} - i & \text{if } i > \frac{h_{\text{len}}}{2} \end{cases}$$

$$\text{and} \quad h_{len} = \frac{history_{length}}{\#idcs_p}$$

- Nowcasting: Concurrent breath phase estimation at the current time point including the $\#idcs_p$ by calculating a nowcast mean using a weighted sum of predicted future values at respective index estimations in the past f_{len} amount of windows:

$$nowcast_{mean} = \frac{\sum_{i=h_{len}}^{f_{\text{len}}} w_i \cdot x_i}{\sum_{i=h_{len}}^{f_{\text{len}}} w_i} \tag{3}$$

$$\text{with} \quad w_i = f_{len} - i \quad \text{and} \quad f_{len} = \frac{forecast_{length}}{\#idcs_p}$$

- Forecasting: Future breath phase prediction by extracting the last index of the window subtracting $\#idcs_p$:

$$forecast = x_{t_{\text{len}} - \#idcs_p} \tag{4}$$
$$\text{with } t_{len} = \frac{tensor_{length}}{\#idcs_p}$$

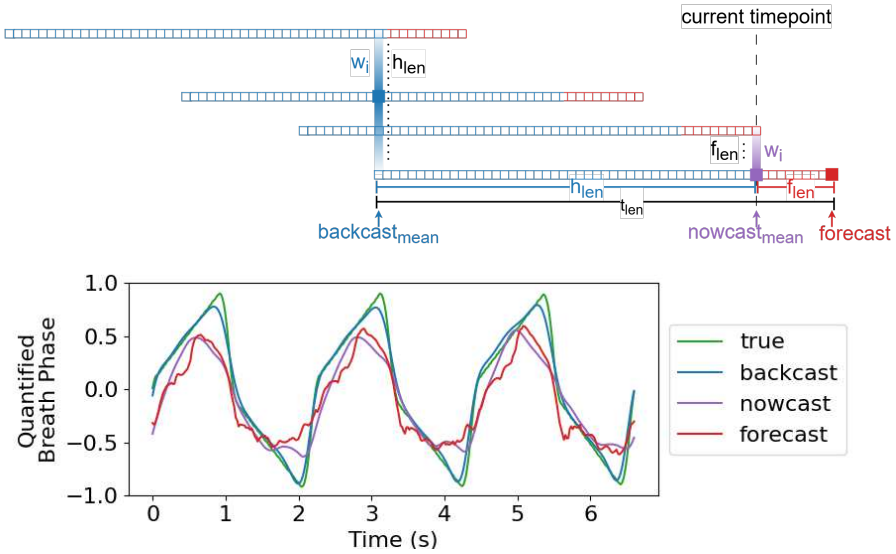


Fig. 2 Casting signal constructions – Continuous signal reconstruction method for backcast (eq. (2)), nowcast (eq. (3)) and forecast (eq. (4)) conditions. Tensors are illustrated with history estimated samples in blue squares and forecast predicted samples in red squares with the latest tensor at the bottom. An example plot with resulting quantified breath phase casting signals is presented underneath.

Each hale (i.e. either inhale or exhale) was annotated with one out of four regularity labels regarding the runners' breathing attachment to the desired LRC to evaluate BPqu regarding the breathing regularity (see Figure 3):

- LRC attached, i.e., breathing is fully coupled to the strides within a deviation tolerance of 5% of the desired LRC:

$$|LRC_i - LRC^*| < 0.05 \cdot LRC^* \tag{5}$$

- In LRC range, i.e., breathing is nearly coupled to the strides within a deviation tolerance of 10% of the desired LRC:

$$0.05 \cdot LRC^* < |LRC_i - LRC^*| < 0.1 \cdot LRC^* \tag{6}$$

- Off LRC range, i.e., breathing is not coupled to strides with a deviation of more than 10% of the desired LRC:

$$|LRC_i - LRC^*| > 0.1 \cdot LRC^* \tag{7}$$

- Anomalous, i.e., the current breath cycle length is less than half or more than double of the average breath cycle length within the past minute:

$$\neg(0.5 < \frac{breath_{len_i}}{\langle breath_{len} \rangle_{1min}} < 2) \tag{8}$$

where $LRC_i = \frac{\langle SR \rangle}{BR_i}$ is the actual LRC, $LRC^* = \frac{\langle SR \rangle}{\langle BR \rangle}$ the desired LRC with averaged $\langle SR \rangle$ and $\langle BR \rangle$ over five cycles, respectively, and the concurrent BR_i. The $breath_{len_i}$ is the length of the concurrent breath and $\langle breath_{len} \rangle_{1min}$ the average breath length of the past minute. The annotation definitions were derived from Harbour et al. Harbour et al. (2023) and Vlemincx et al. Vlemincx et al. (2013b). The breath regularity labeling resulted in 49.3% LRC attached, 21.2% in LRC range, 21.6% off LRC range and

7.9% anomalous breathing annotations. Finally after evaluation, a quantized model is trained with the whole dataset for deployment.

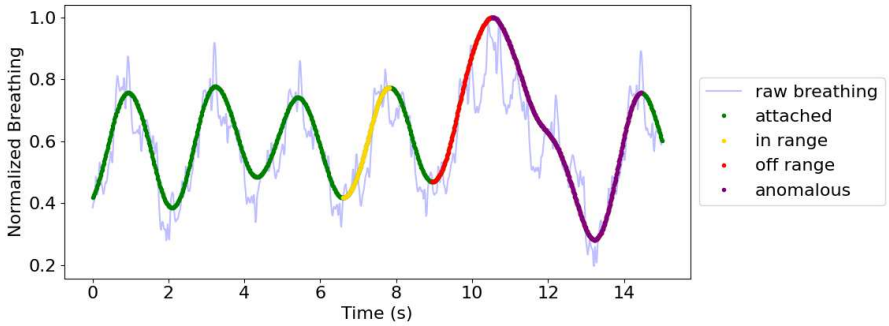


Fig. 3 Breathing regularity annotations – Labelled breathing sequence according to the determined breathing regularity annotations: LRC attached (eq. (5)), in LRC range (eq. (6)), off LRC range (eq. (7)), and anomalous (eq. (8)).

3.5.3 Metrics

A model configuration with an optimal complexity-accuracy trade-off is identified through the hyper-opt procedure. The complexity is measured in the number of model parameters (#params) and accuracy is measured in the nRMSE of the actual and estimated surrogate forecast tensors. The euclidean distance (d) of the nRMSE and logarithmic #params was calculated to reveal a decisive complexity-accuracy trade-off value. Following categories were experimentally investigated to optimize the BPqu architecture in the complexity-accuracy trade-off:

1. Input sensor source: Breathing while applying LRC inherits a dependency between the breathing and stride signals. Therefore, the complexity-accuracy trade-off is examined when feeding with or without adding the stride information: (a) only breathing data (CAP); (b) breathing & stride data (MSF).
2. Forecast length: The BPqu output tensor consists of a history and forecast tensor. The ratio between history and forecast tensor was kept consistently to 5:1

(history:forecast), i.e., the history tensor increased proportionally to the forecast tensor. The forecast tensor length was changed to match the breathing within following windows: (a) one step with $step_{len_M} = 379ms$ (STEP); (b) one hale with $\frac{breath_{len_M}}{2} = 936ms$ (HALE); (c) one breath with $breath_{len_M} = 1.872s$ (BREATH).

3. CNN number of layers (CNN #L): Different layering strategies were applied with different shapes, however, the maximum hidden size was kept consistent: (a) one hidden layer; (b) three hidden layers in a triangularly growing shape; (c) five hidden layers in an asymmetric diamond shape with a growing and differently decreasing side, respectively.
4. CNN hidden size and kernel size factors (CNN HF and KF): The number of nodes in the hidden layers and kernel sizes were multiplied by following factors compared to the baseline, respectively: (a) 0.125; (b) 0.25; (c) 0.5; (d) 1.0.
5. LSTM number of layers and hidden size (LSTM #L and HF): The number of layers and the predefined hidden size of the LSTM was alternated by identical factors defined in 3) and 4).

The finally revealed BPqu architecture with an optimal complexity-accuracy trade-off is used for further model evaluation.

A SHAP (SHapley Additive exPlanations) Lundberg and Lee (2017) analysis is conducted to explain the model performance dependency on breathing and stride data via the nRMSE based on the hyper-opt output for varying input data: no breathing or stride data (e_0); only breathing data (e_b); only stride data (e_s); both breathing and stride data (e_{bs}).

The model performance depending on the four breathing rhythm regularities is evaluated in the three casting conditions. The BPqu output is evaluated with three metrics:

- Hale change event detection (eq. (9)): Precision (P), recall (R) and the F1-score (F1) of detected hale change events (i.e. timepoints between inhalation and exhalation phases) are determined based on a peak detection (*scipy1.15.2* Virtanen et al. (2020)) in the reconstructed casting event surrogate signals (see blue signal in Fig. 1). The hale timepoints are classified as true positive (T_{pos}) if there are matching local maxima or minima in the true and predicted event surrogate signals within an event centered two-step length tolerance window. The optimal prominence was determined by identifying the maximum F1-score in the nowcast condition by creating a precision-recall curve (PRC) iterating over prominence values ranging from 0.1 to 0.9 of the true surrogate peak and valley amplitude values (maximal peak/valley at ±128 for int8-quantized values). Otherwise, wrongly identified hale changes in prediction signals are classified as false positives (F_{pos}) and wrongly missed hale changes are classified as false negatives (F_{neg}). Additionally, the temporal deviation of T_{pos} hale change timepoints is investigated in mean (M) and root mean squared error (RMSE) in ms.

$$\begin{aligned} \mathrm{P} = \frac{T_{\mathrm{pos}}}{T_{\mathrm{pos}} + F_{\mathrm{pos}}}, \quad \mathrm{R} = \frac{T_{\mathrm{pos}}}{T_{\mathrm{pos}} + F_{\mathrm{neg}}}, \quad \mathrm{F1} = 2 \cdot \frac{\mathrm{P} \cdot \mathrm{R}}{\mathrm{P} + \mathrm{R}} \\ \textit{and} \quad \mathrm{RMSE}_{T_{\mathrm{pos}}} = \sqrt{\frac{1}{|T_{\mathrm{pos}}|} \sum_{i \in T_{\mathrm{pos}}} e_i^2} \\ \textit{with} \quad e_i = \hat{x}_i - x_i \end{aligned} \tag{9}$$

where e_i is the error for true positive samples (T_{pos}), x_i the true hale change timepoint, $\hat{x}_i$ the estimated hale change timepoint, and $|T_{\mathrm{pos}}|$ the number of true positive samples.

- Hale phase determination (eq. (10)): Depending on the detected hale change timepoints, distinct hale phases are identified and the accuracy (Acc) is calculated for correctly identified inhalation (T_{in}) and exhalation (T_{ex}) phases compared

to misclassified inhalation (F_{in}) and exhalation (F_{ex}) phases. The breath cycle phase estimation deviation is likewise measured on correctly estimated hale phases ($T_{in} \cup T_{ex}$) by evaluating the M phase error e_i and the RMSE of the estimated base surrogate signal (see red signal in Fig. 1) in conversion to radians (rad; one hale between 0 and 1 equals 0 and π) in correctly classified hale phases:

$$\begin{aligned} \mathrm{Acc} &= \frac{\mathrm{T_{in}} + \mathrm{T_{ex}}}{\mathrm{T_{in}} + \mathrm{F_{in}} + \mathrm{F_{ex}} + \mathrm{T_{ex}}} \\ \text{and} \quad \mathrm{RMSE_T} &= \sqrt{\frac{1}{|T_{\mathrm{in}} \cup T_{\mathrm{ex}}|} \sum_{i \in \mathrm{T_{in}} \cup \mathrm{T_{ex}}} e_i^2} \\ \text{with} \quad e_i &= \begin{cases} y_i - \hat{y}_i & \text{if inhale} \\ \hat{y}_i - y_i & \text{if exhale} \end{cases} \end{aligned} \tag{10}$$

where e_i is the error for true positive samples ($T_{\mathrm{in}} \cup T_{\mathrm{ex}}$), y_i the true hale phase value, $\hat{y}_i$ the estimated hale phase value, and $|T_{\mathrm{in}} \cup T_{\mathrm{ex}}|$ the number of true positive samples.

- Breath phase quantification (eq. (11)): The resulting breath phase quantification output (see yellow signal in Fig. 1) reconstructed from the continuous event and base surrogate signals is evaluated by computing the nRMSE over all samples independent from hale change point detection and hale phase classification outputs for true and estimated quantified breath phase values z_i and $\hat{z}_i$, respectively:

$$\begin{aligned} \mathrm{nRMSE} &= \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (z_i - \hat{z}_i)^2}}{z_{max} - z_{min}} \\ \text{with} \quad & z_{max} - z_{min} = 2 \\ \text{for} \quad z_{max} = 1 \quad & \text{and} \quad z_{min} = -1 \end{aligned} \tag{11}$$

The metrics were chosen to minimize time in incorrectly detected hale phases, which can arise from false positives, false negatives, or temporal deviations that distort phase boundaries. Thus, F1-score and $RMSE_T$ are appropriate, as feasibility depends on their combination and the hale length. Detected hale frequency is further affected by imbalanced errors, which is minimized by selecting the threshold where precision equals recall (see Fig. 6).

Considering the distribution of these metrics for an ideally shaped estimation signal with an average hale length of two and a half step lengths l, the expected metric values can be characterized for the following extreme cases dependent in the phase shift φ: (a) perfect phase detection ($\varphi = 0 \cdot l$), (b) random phase assignment uncorrelated with the true phase ($\varphi = \sqrt{\frac{1}{l}E(x)} = \frac{l}{\sqrt{3}}$), (c) a phase shift of one step length (worst-case detected temporal offset; $\varphi = l$), and (d) a 180-degree phase shift (complete phase inversion; $\varphi = 2.5 \cdot l$). Table 1 presents the expected metric values.

Table 1 Expected classification metric and error measure values for extreme cases with (a) zero, (b) random, (c) one step length, (d) inverse phase shifts.

	P (%)	R (%)	F1 (%)	Acc (%)	$\text{RMSE}_{\text{T}_{\text{pos}}}$ (ms)	RMSE_{T} (rad)	nRMSE
(a)	100	100	100	100	0	0	0
(b)	40	40	40	40	218.8	0.158π	0.382
(c)	100	100	100	100	379.0	0.292π	0.360
(d)	0	0	0	0	–	–	0.577

Given a similarly shaped estimation signal, the nRMSE is a monotonic function of the phase shift between the true and the estimated signal. While the other metrics focus on the accuracy of the applied detection method by peak threshold, a low nRMSE could indicate the possibility of improved detection heuristics tailored to a specific guidance application. Relating to the average BR of 32.1 bpm, an nRMSE of 1%,

i.e., a breath phase deviation of 0.02π, is representative for a temporal deviation of approximately 18.7 ms in the breath phase quantification.

4 Results

BPqu is analysed regarding its trade-offs between complexity and accuracy as well as latency and accuracy in the following subsections. Thereby, the impact of selected sensor sources, forecast lengths, CNN architecture (section 4.1), and breathing regularity in different casting conditions (section 4.2) is investigated.

4.1 Sensor source, forecast length & CNN architecture

Table 2 and Figure 4 presents the hyper-opt results. Averaging the Euclidean distance metrics d for various forecast lengths (BREATH, HALE, STEP) and sensor sources (CAP, MSF) revealed that STEP forecast length provides the optimal complexity-accuracy trade-off. Specifically, the STEP forecast length resulted in $d_{STEP} = 11.12$, which was lower compared to HALE ($d_{HALE} = 12.91$) and BREATH ($d_{BREATH} = 14.09$) forecast lengths. On average, using CAP ($d_{CAP} = 12.57$) as only sensor source provides a better complexity-accuracy trade-off than applying MSF ($d_{MSF} = 12.85$).

Focusing on averaged model sizes, forecast length STEP ($\#Params_{STEP} = 6.9e^4$) provides the smallest model sizes followed by HALE ($\#Params_{HALE} = 4.1e^5$) and BREATH ($\#Params_{BREATH} = 1.3e^6$). Using only CAP ($\#Params_{CAP} = 5.7e^5$) as sensor source reveals lower model sizes than applying MSF ($\#Params_{MSF} = 6.3e^5$).

Averaging the accuracies, HALE ($nRMSE_{HALE} = 0.135$) reveals the best accuracy followed by STEP ($nRMSE_{STEP} = 0.137$) and BREATH ($nRMSE_{BREATH} = 0.155$). Using the forecast length STEP and applying MSF ($nRMSE_{MSF|STEP} = 0.125$) is more accurate than using CAP ($nRMSE_{CAP|MSF} = 0.149$). If forecast lengths HALE or BREATH are applied, the accuracy is better by using

only CAP ($nRMSE_{CAP|HALE} = 0.130, nRMSE_{CAP|BREATH} = 0.143$) instead of MSF ($nRMSE_{MSF|HALE} = 0.139, nRMSE_{MSF|BREATH} = 0.167$).

Consequently, $BPqu_{CAP|STEP}$ with a one-layer LSTM with one-quarter of the original hidden-size and a five-layer CNN with one-eigth of the orginal hidden and kernel size is revealed as the model with the optimal complexity-accuracy trade-off. Nevertheless, the most accurate model is $BPqu_{MSF|STEP}$ with a three-layer LSTM with half of the original hidden-size and a five-layer CNN with one-quarter and one-eigth of the original hidden and kernel sizes, respectively. $BPqu_{MSF|STEP}$ contains a model size of 84,528 trainable parameters that requires 330.19kB memory storage and deployability tests revealed a model execution processing delay of $14ms$ ($8ms$ model inference including $1ms$ pre-quantization and $5ms$ breath phase reconstruction). Consequently, $BPqu_{MSF|STEP}$ was selected for further application because the required memory storage is available on the aimed hardware and the processing delay is less than the sampling delay of $20ms$ on the MCU with a sampling rate of $50Hz$ that enables a model execution with each incoming new sample.

Conclusively, $BPqu_{MSF|STEP}$ takes two one-dimensional signals (breathing and steps) with the length of five steps into a sequential model with a three-layer LSTM with a maximum hidden size of 16 and a five-layer CNN with a maximum hidden size of 8 and kernel size of 4. Figure 5 presents the BPqu architecture and one example sample including the true signals and estimated surrogate and quantified breath phase signals based on the breathing raw and acceleration data within one BPqu tensor. The breathing regularity and casting evaluations concerning hale exchange time point estimation, distinct breath phase accuracy and continuous breath phase quantification performance is rendered with $BPqu_{MSF|STEP}$ (named BPqu in the remaining article).

The SHAP analysis showed that BPqu achieved the lowest error when both breathing and stride data were provided ($e_{bs} = 0.101$) followed by BPqu with only breathing data ($e_b = 0.107$). Performance was equally poor without breathing data, regardless

of whether stride data alone or no data were provided ($e_s = e_0 = 0.227$). The SHAPley values, based on these error metrics, are $\phi_b = -0.123$ for the breathing data and $\phi_s = -0.003$ for the stride data channel.

Table 2 Model architectures of sensor source and forecast length combinations with the best complexity-accuracy trade-off.

	CAP BREATH	CAP HALE	CAP STEP	MSF BREATH	MSF HALE	MSF STEP
LSTM #L	3	1	1	1	1	3
LSTM HF	0.50	1.00	0.25	0.25	0.50	0.50
CNN #L	5	5	5	5	5	5
CNN HF	0.125	0.125	0.125	0.125	0.250	0.250
CNN KF	1.000	1.000	0.125	0.250	0.250	0.125
nRMSE	0.143	0.130	0.149	0.167	0.139	0.125
#Params	$1.31e^6$	$3.33e^5$	$5.37e^4$	$1.31e^6$	$4.89e^5$	$8.45e^4$
Eucl.D.	14.09	12.72	10.89	14.09	13.10	11.35

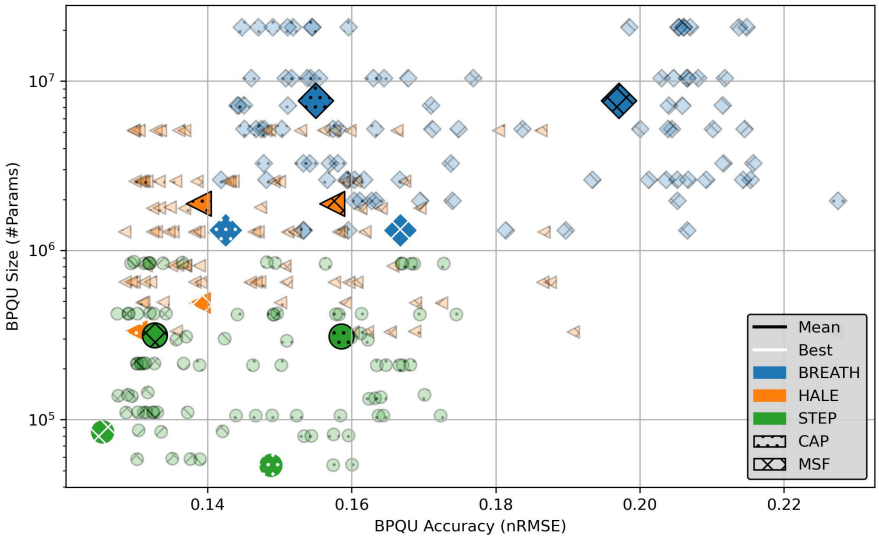


Fig. 4 Complexity-accuracy trade-off evaluation – Scatter plotted BPqu model combinations regarding their size in #Params and accuracy in nRMSE. The best BPqu combinations are highlighted for each sensor source and forecast length combination.

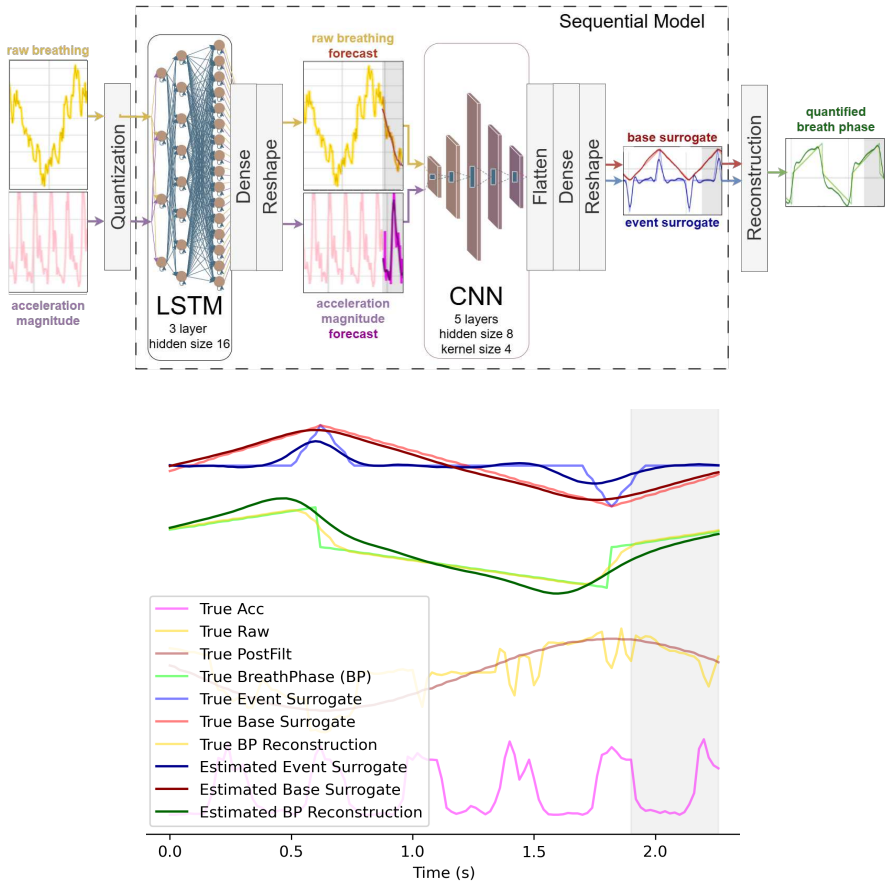


Fig. 5 Model architecture – BPqu model architecture with an example sample illustrating the history window (non-shadowed) and forecast window (gray-shadowed) including true and estimated signals.

4.2 Breathing regularity & casting

Table 3 concludes the BPqu accuracy performance metrics in the back-, now- and forecasting conditions.

BPqu detects hale change events precisely in backcasting conditions ($F1_{back}$ = 98.0%) with a temporal deviation of −1.0ms and an averaged RMSE of 39.7ms. The performance is decreased in now- and forecasting conditions ($F1_{now}$ = 93.1% with Δt_{now} = −7.2ms and $RMSE_{now}$ = 128.0ms and $F1_{fore}$ = 88.5% with Δt_{fore} = −6.6ms and $RMSE_{fore}$ = 138.1ms).

Table 3 Casting performance metrics of the hale change timepoint detection in precision (P), recall (R), F1-score (F1), average temporal deviation and error ($\Delta tM \pm RMSE$) in ms. In addition, the inhalation and exhalation phase determination accuracy (Acc) including average phase deviation and error ($\Delta \phi M \pm RMSE$) in radians (rad). Finally, the quantified breath phase reconstruction performance ($\Delta nRMSE$) is stated in percent (%).

	P	R	F1	Acc	Δt M±RMSE	$\Delta \phi$ M±RMSE	$\Delta BPqu$ nRMSE
back	96.6	99.6	98.0	96.4	-1.0±39.7	0.01±0.37π	0.085
now	92.6	93.6	93.1	85.6	-7.2±128.0	0.03±0.34π	0.185
fore	85.1	92.3	88.5	83.0	-6.6±138.1	0.04±0.33π	0.231

Figure 6 presents the PRC for the nowcast hale change event detection threshold determination. The PRC maximum reveals that a threshold of 0.3 of the maximum amplitude of the event surrogate (prominence equals 38.4 for int8-quantized data) is optimal with a nowcast F1-score is 93.1% with the most balanced precision of 92.6% and recall of 93.6% ratio. In addition, the hale change F1-score of different casting conditions and breathing regularity levels is presented in a heatmap including the barplotted temporal mean deviation and RMSE in ms. The heatmap reveals that hale change event detection performance correlates with breathing regularity. The backcasting F1-score varies from 89% to 99%, nowcasting from 77% to 96% and forecasting between 70% and 92%. Breathing regularity also influences the temporal mean deviation and the RMSE of the hale change timepoint detection. Whereas the mean temporal deviation in backcasting is stable independently from breathing regularity within a range of -2 to $+2$ ms, the RMSE increases from 28 to 75 ms with increasing breathing irregularities. In now- and forecasting conditions, the temporal deviation lag as well as the RMSE increases with enhanced breath irregularities. Δt_{now} varies from -3 ± 119 to 35 ± 152ms and Δt_{fore} ranges between -6 ± 125 to $+36 \pm 173$ms.

Table 3 additionally presents the BPqu accuracy in determining distinct breath phases, i.e. inhalation and exhalation phases, and reveals an accuracy of 96.4% in backcasting with a continuous hale phase deviation and error of $0.01 \pm 0.37\pi$. Accuracy

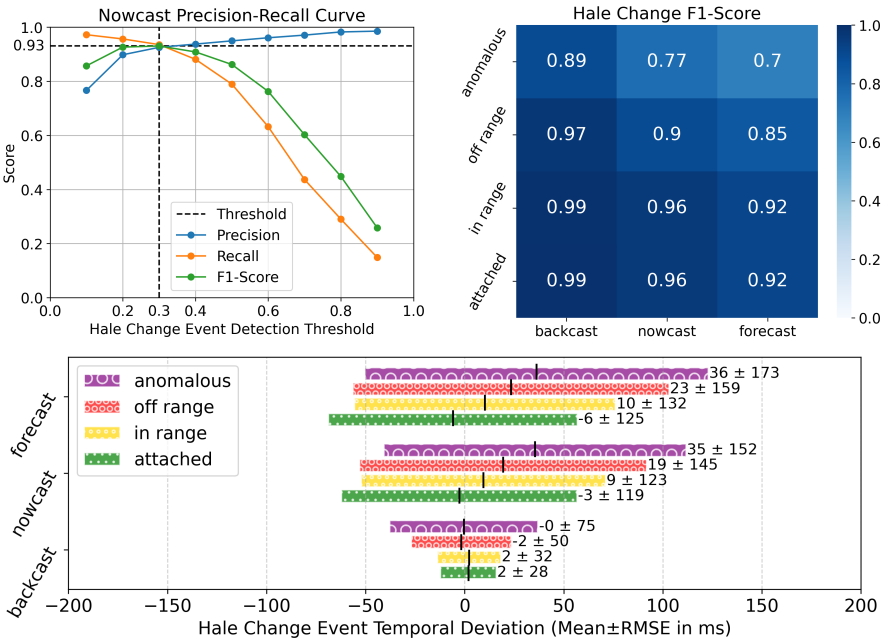


Fig. 6 Hale change event detection evaluation – Nowcast hale change precision-recall curve illustrating the maximum F1-score for hale change minima and maxima detection threshold determination. Hale change F1-score heatmap presenting hale change timepoint detection F1-score including barplotted temporal deviation and error over different breath phase regularity levels (LRC attached, in LRC range, out of LRC range, anomalous) grouped in casting conditions (backcast, nowcast, forecast).

performance metrics are decreased in now- and forecasting ($Acc_{now} = 85.6\%$ with $\Delta\phi_{now} = 0.03 \pm 0.34\pi$ and $Acc_{fore} = 83.0\%$ with $\Delta\phi_{fore} = 0.04 \pm 0.33\pi$ phase deviation, respectively).

Figure 7 presents the nowcast hale phase detection accuracy confusion matrix illustrating that inhalation (86%) and exhalation (85%) phases are almost equally precise estimated. In addition, the hale phase accuracy heatmap reveals the hale detection accuracy performance over differnt casting conditions and breathing regularities. The hale phase detection accuracy correlates with the breathing regularity. Depending on the breathing regularity, the accuracy varies from 87% to 98% in backcasting, from 69% to 90% in nowcasting and between 64% and 88% in forecasting. Whereas the breath phase estimation RMSE correlates with breathing irregularity, the mean phase deviation lag is almost consistent over all breathing regularites. The mean deviation

lag and RMSE varies between $-0.017 \pm 0.347\pi$ and $0.020 \pm 0.398\pi$ in backcasting, from $-0.002 \pm 0.323\pi$ to $0.035 \pm 0.376\pi$ in nowcasting and between $-0.002 \pm 0.314\pi$ and $0.049 \pm 0.355\pi$ in forecasting.

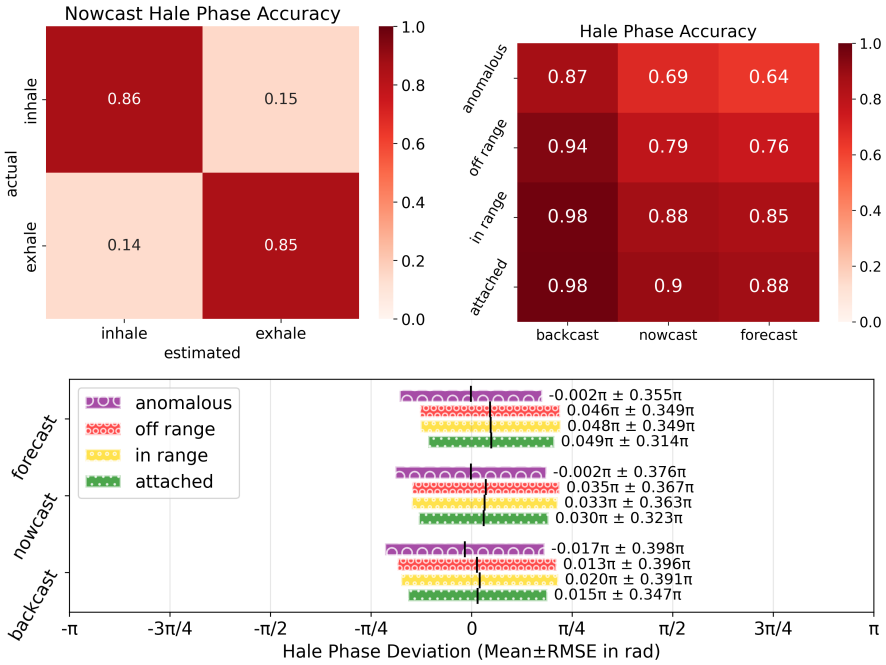


Fig. 7 Hale phase determination evaluation – Nowcast hale phase confusion matrix presenting inhalation and exhalation phase determination accuracy. Hale phase accuracy heatmap illustrating hale phase classification accuracies including bar plotted hale phase deviation and error over different breath phase regularity levels (LRC attached, in LRC range, out of LRC range, anomalous) grouped in casting conditions (backcast, nowcast, forecast).

Table 3 conclusively presents the performance of BPqu when a quantified breath phase is reconstructed from the two surrogate signals. The nRMSE for back-, now- and forecasting amounts 0.085, 0.185, and 0.231, respectively, that reveals a BPqu accuracy of 91.5%, 81.5%, and 76.9%, respectively. Assuming that 1% corresponds to 18.7 ms (see section 3.5), temporal deviations in breath phase estimation can be approximated as 159 ms, 346 ms, and 432 ms, respectively.

Figure 8 presents the performance of BPqu when a quantified breath phase is reconstructed from the two surrogate signals. The plotted nowcast breath phase quantification error presents the error deviation over a normalized hale averaged over inhalation and exhalation phases including the average nRMSE in each hale quartile. An increasing nRMSE is observed over the course of a hale, based on average values computed for each quarter of its duration: 0.126, 0.135, 0.157, and 0.271 (approx. 236ms, 252ms, 294ms, and 508ms). In addition, the breath phase quantification heatmap illustrates that the nRMSE correlates with breathing irregularity and varies between 0.05 and 0.21 (approx. 94 ms and 393 ms) in backcasting for different breathing regularities. The performance is reduced in now- and forecasting and the nRMSE varies from 0.15 to 0.31 (approx. 281 ms and 580 ms) and between 0.19 and 0.37 (approx. 355 ms and 692 ms), respectively.

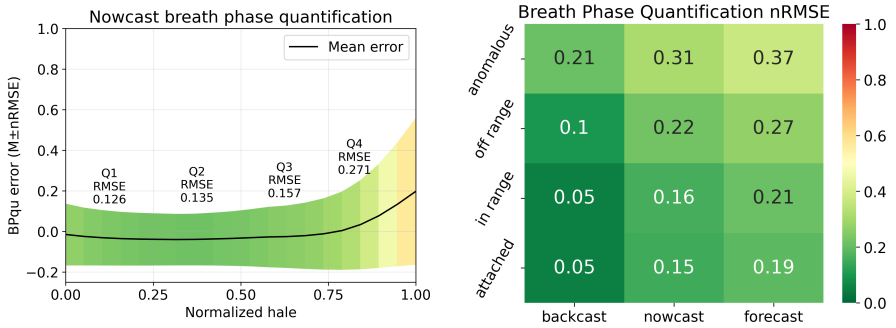


Fig. 8 Breath phase quantification evaluation – Plotted nowcast breath phase quantification normalized mean error and nRMSE averaged over normalized inhalation and exhalation phases. Heatmap of the reconstructed quantified breath phase nRMSE over different breath phase regularity levels (LRC attached, in LRC range, out of LRC range, anomalous) grouped in casting conditions (backcast, nowcast, forecast)

5 Discussion

The aim of BPqu is to generate new sonified breathing guidance opportunities in wearable breathing sensing devices with real-time feedback interfaces. The hyperopt strategy of selecting optimal model architecture parameters via computing the

euclidean distance from nRMSE and logarithmic #Params turned out to be a successful approach to reduce model complexity with maintaining the model accuracy performance to achieve the optimal complexity-accuracy trade-off. The hyper-opt revealed that the shorter the tensor lengths the more effective is sensor fusion of breathing and stride data. The stride data input mostly supports the motion artifact removal component in the CNN which is especially important if not the whole breathing cycle is fed into the model as given in the STEP forecast length scenario. Nevertheless, data-level sensor fusion causes more amount of data processing, i.e. a more complex size measured in #Params, whereby decreasing the hidden size can counteract to maintain an optimal complexity-accuracy trade-off.

BPqu is evaluated based on its complexity-accuracy trade-off that is optimized by using the euclidean distance between model size and accuracy as loss function in the hyper-opt procedure that was limited to the investigation of the number of layers, hidden size and kernel size. In the future, the complexity-accuracy trade-off could be improved in the hyper-opt by enhancing sparsity through pruning and efficiency via dilation scaling. Furthermore, weights could be integrated into the euclidean distance loss function if accepted feedback accuracy errors and allowed model sizes for deployment are foreknown.

BPqu is investigated in its surrogate components and the reconstructed quantified breath phase output. Depending on the breathing guidance objective, the necessary BPqu output metrics can be selected between hale change timepoints from event surrogate signal, hale phase from base surrogate signal or quantified breath phase from reconstruction signal. BPqu shows potential to satisfy real-time breathing guidance constraints, as zero-delay hale change events are detected with a F1-score of 93.1%, and the RMSE remains below 152 ms even under anomalous breathing patterns. An advantage of the architecture is that the input signal size of a length of five steps is relatively short, which enables fast adaption to breathing pattern changes. However, in

severe cases, correctly detected hale change events may be mistakenly assigned to the wrong step cycle, potentially leading to user discomfort due to desynchronized feedback. These issues must be evaluated under real-time, in-field conditions to assess user acceptance of BPqu-generated breathing guidance. To improve reliability, on-device safeguards, e.g. watchdog timers, could be employed to detect delayed or invalid model output. In such cases, the system could fall back to a simple, rhythmic breathing pattern to maintain consistent and uninterrupted feedback. Furthermore, future work can explore the integration of confidence threshold mechanisms to reduce false positives in critical transitions and enhance decision safety. Finally, a comparative study could be conducted to evaluate the BPqu-based breathing guidance system against existing approaches, e.g. *RunRhythm* Harbour et al. (2024) and *Strive* Valsted et al. (2017), in terms of latency, accuracy, and user preference.

The quantified breath phase can be evaluated regardless of the hale change event detection and hale phase determination. However, the output metric is normalized within the maximum range of the quantified breath phase. It can not be directly interpreted as breath phase in radians or temporal deviation in ms, although an approximation using the mean BR suggests a deviation of 346 ms at zero delay, averaged across all breathing regularities. Nevertheless, the breath phase quantification is a proper approximation to the real breath cycle during hale phases, but weak estimation around hale change events must be considered. Therefore, the quantified breath phase output should be manipulated if a hale change timepoint is predicted to ensure a sonification that does not interfere with the users' perception.

BPqu is proven to be applicable for real-time breath phase estimation and is designed to run on a preselected MCU. However, the MCU cannot run RNNs which limits the efficiency of the model because no TFLM inference is enabled and TensorFlow H5 models are required. As TensorFlow-Lite is particularly developed for edge-AI use cases, the MCU selection should be revised or MCU developers are

requested to enable TFLM inference also for RNNs. Additionally, modifying the network architecture to remove the LSTM dependency, potentially using lightweight temporal convolution or attention-based modules, could solve the efficiency limitation. However, a quantitative comparison with a TFLM-compatible pipeline is needed to assess trade-offs in latency, size, and accuracy for optimizing BPqu on low-power edge devices.

BPqu, implemented with deep learning and offering limited inherent explainability, was analyzed using SHAP, which confirmed that combining breathing and stride data best supports artifact removal and accurate breath phase estimation. The current architecture places the LSTM before the CNN to derive surrogate functions from temporally forecasted signals, but alternative LSTM placements (before, between, or after CNN layers) should be investigated to evaluate their impact on accuracy and latency.

BPqu was trained and evaluated using breath data from the SRS Bernhart et al. (2023), making it specific to a single sensor device. However, its evaluation does not serve to assess the measurement accuracy of the SRS. The BPqu evaluation assumes breath data without temporal deviations, meaning the results are applicable to any breath data but do not validate the accuracy of the breathing sensor itself. Consequently, BPqu can be utilized with breath data from any sensor, provided it is trained on a dataset specific to that sensor. In the future, training BPqu on breath data from multiple sensor sources could enhance its applicability. While the proposed architecture cannot be applied to additional sensory data without modification, the main improvement, the surrogate signals, is located on the output side of the architecture, and, therefore, easily applicable to arbitrary input by just extending input layers.

BPqu requires training outside of the MCU and is biased on balanced breath duty cycle data coming from nineteen young female healthy runners. Data including unbalanced breath duty cycles from a more demographically diverse group of participants

could be trained offline via generative AI using data augmentation. In addition, online learning should be enabled via tinyML methods to improve the BPqu applicability to different regular breathing patterns with asymmetric duty cycles and broader breath frequency ranges as well as more irregular breath patterns (e.g. sighs, yawns, coughs, etc.).

BPqu shows competitive, context-specific performance across backcasting, nowcasting, and forecasting, extending prior ML/DL-based respiratory analysis. In offline processing scenarios, an F1-score of 92% was reported for normal breathing waveform classification McClure et al. (2020), and 91% agreement was achieved for distinct breath phase determination (inspiration and expiration) Jácome et al. (2019). This compares to BPqu's backcasting performance, which yielded an F1-score of 98.0% for hale change event detection, 96.4% accuracy in discrete hale phase classification, and 91.5% in continuous breath phase estimation. A breathing waveform forecasting algorithm for a 300 ms window reported an nRMSE of 0.11 for regular breathing and 0.56 for irregular breathing, outperforming standard filters by 5% Murphy and Pokhrel (2009). Compared to BPqu's forecasting for one step (379 ms), with an nRMSE of 0.19 for regular and 0.37 for irregular breathing, BPqu does not reach the prior performance in regular breathing prediction but demonstrates higher accuracy for irregular patterns. Additionally, a low-latency breath phase estimation approach (<10 ms delay) achieving <20° deviation for over 90% of the data in the 8–12 Hz frequency range (brain waves instead of respiratory signals) was recently reported using a reduced-complexity TCN with 70,000 trainable parameters, outperforming both a complex finite impulse response (cFIR) filter and a Kalman filter-based state-space estimation. The TCN performance is comparable to BPqu's nowcasting performance, which achieves a breath phase deviation of $5.4° \pm 61.2°$ for 85.5% of the data within a lower frequency range (0.4–0.8 Hz), using a model with 84,528 trainable parameters. These results highlight BPqu's suitability for real-time breath phase quantification, matching

state-of-the-art ML/DL methods, while conventional FIR, Kalman, and DAMOCO approaches remain less robust under irregular breathing.

BPqu enables innovative breath phase based auditory guidance in running. Different auditory breathing guidance strategies will be investigated in future studies. Although BPqu cannot provide guidance based on breath volume parameters, rhythmic-controlled strategies (e.g., LRC-based) can be instructed without lagging delay with BPqu even if breathing suddenly changes. Observational breathing guidance can be improved by incorporating the concurrent breath phase and hale change events through BPqu to guide runners' hale phases, e.g., prolonging concurrent exhalation phases. Therefore, BPqu has potential to guide runners towards a rhythmic-controlled breathing to compensate respiratory distress, overcome cognitive-emotionally barriers during sports activities, and enhance running enjoyment towards improved well-being.

Assuming a successful in-field evaluation, the proposed system may offer potential for integration with commercial wearable platforms or smart textiles, highlighting opportunities for collaboration with industry stakeholders in the sports and health technology sectors. Future work could also explore market potential for this type of edge-AI-guided feedback system in consumer fitness and recreational running applications.

6 Conclusion

This work presents the development and evaluation of BPqu that is a breath phase quantifier deep learning algorithm for real-time application on edge. The algorithm aims to enable new opportunities in breathing guidance in running by predicting breathing and establishing breath phase information in zero-delay real-time. BPqu is implemented and evaluated based on a laboratory dataset. The Bpqu performance is evaluated in backcasting, nowcasting, and forecasting conditions and the influence

of different breathing regularities is investigated. The model is optimized to fulfill complexity-accuracy trade-off requirements to match edge-AI requirements while maintaining accuracy performance.

Key results are that short forecasting periods are beneficial to fulfill complexity-accuracy trade-off requirements and sensor fusion is especially successful in short forecasting periods to accomplish real-time motion artifact removal. The approach to split a signal to be estimated into the sum of surrogate signals which are easier to predict, seems promising and will be investigated further in future work. In addition, hale change events can be estimated correctly with a F1-score of 93.1% with a temporal deviation lag of -7.2 ± 128.0ms with zero processing delay. Moreover, inhalation and exhalation phases are determined correctly with an accuracy of 85.6% and the continuous hale phase is determined with a phase deviation of $0.03 \pm 0.34\pi$ with zero processing delay. Independently from breath events or phases, BPqu quantifies the concurrent breath phase with an accuracy of 81.5% within zero delay.

In summary, BPqu enables concurrent continuous precise breath phase determination in real-time and can enhance new breathing guidance opportunities. BPqu is deployed on a MCU that is integrated in a custom-developed textile wearable stride and breathing sensor. Thus, future field studies are enabled to implement and test concurrent breathing guidance strategies. Users' feedback will provide insights if BPqu's accuracy is sufficient to provide intuitive and not distracting breathing guidance in real-time. BPqu offers the potential to introduce novel modalities for addressing cognitive-emotional aspects, thereby enhancing motivation in sports participation and improving individual performance.

Declarations

- Funding: blinded

- Conflict of interest/Competing interests: The authors have no competing interests to declare that are relevant to the content of this article.
- Ethics approval and consent to participate: blinded
- Consent for publication: Written informed consent was obtained from all participants.
- Data availability: Can be discussed after publication.
- Code availability: Can be discussed after publication.
- Author contribution: blinded

References

AB I (2024) Imagimob AI - Edge AI Development Platform. URL https://www.imagimob.com, accessed: 2024-01-26

Aliverti A (2016) The respiratory muscles during exercise. Breathe 12(2):165–168. https://doi.org/10.1183/20734735.008116

Archiza B, Leahy MG, Kipp S, et al (2021) An integrative approach to the pulmonary physiology of exercise: when does biological sex matter? European Journal of Applied Physiology 121(9):2377–2391

Banbury C, Zhou C, Fedorov I, et al (2021) Micronets: Neural network architectures for deploying tinyml applications on commodity microcontrollers. In: Smola A, Dimakis A, Stoica I (eds) Proceedings of Machine Learning and Systems, pp 517–532

Bernhart S, Harbour E, Kranzinger S, et al (2023) Wearable chest sensor for stride and respiration detection during running. Sports Engineering 26(1):19. https://doi.org/10.1007/s12283-023-00410-y

Bernhart S, Guth F, Poredos B, et al (2024) Warp knitting and custom chip design for breathing guidance apparels. In: Proceedings of the International Symposium on

Electronic Art (ISEA). Loughborough University, p 0, https://doi.org/10.17028/rd.lboro.27051571.v1

Bernhart S, Finkenzeller T, Kutil R (2025) Breathe-run-guide: A review of personal breathing guidance systems in running. SN Computer Science 6(6):609. https://doi.org/10.1007/s42979-025-04142-7

Bloch S, Lemeignan M, Aguilera-T N (1991) Specific respiratory patterns distinguish among human basic emotions. International Journal of Psychophysiology 11(2):141–154. https://doi.org/https://doi.org/10.1016/0167-8760(91)90006-J

Daley MA, Bramble DM, Carrier DR (2013) Impact loading and locomotor-respiratory coordination significantly influence breathing dynamics in running humans. PLOS ONE 8(8):1–10. https://doi.org/10.1371/journal.pone.0070752

Developers T (2024) TensorFlow: Open Source Machine Learning Framework. URL https://www.tensorflow.org, accessed: 2024-01-26

Ellis C, Masood SZ, Tappen MF, et al (2013) Exploring the trade-off between accuracy and observational latency in action recognition. International Journal of Computer Vision 101(3):420–436. https://doi.org/10.1007/s11263-012-0550-7

Feigenspan A (2017) Atmung und Physiologie der Atemgase, Springer Berlin Heidelberg, Berlin, Heidelberg, chap 4, pp 153–207. https://doi.org/10.1007/978-3-662-54117-3_4

Foussier J, Teichmann D, Jia J, et al (2014) An adaptive kalman filter approach for cardiorespiratory signal extraction and fusion of non-contacting sensors. BMC medical informatics and decision making 14:1–15. https://doi.org/10.1186/1472-6947-14-37

Harbour E, Lasshofer M, Genitrini M, et al (2021) Enhanced breathing pattern detection during running using wearable sensors. Sensors 21(16). https://doi.org/10.3390/s21165606

Harbour E, Stöggl T, Schwameder H, et al (2022) Breath tools: A synthesis of evidence-based breathing strategies to enhance human running. Frontiers in Physiology 13:813243. https://doi.org/10.3389/fphys.2022.813243, eCollection 2022

Harbour E, Rheden Vv, Schwameder H, et al (2023) Step-adaptive sound guidance enhances locomotor-respiratory coupling in novice female runners: A proof-of-concept study. Frontiers in Sports and Active Living 5. https://doi.org/10.3389/fspor.2023.1112663

Harbour E, van Rheden V, Rieser H, et al (2024) Custom smartphone application to guide locomotor-respiratory coupling in the field using step-adaptive breathing sounds. Journal of Visualized Experiments (JoVE) p e66733

Hornik K, Stinchcombe MB, White HL (1989) Multilayer feedforward networks are universal approximators. Neural Networks 2:359–366

Hribernik M, Umek A, Tomažič S, et al (2022) Review of real-time biomechanical feedback systems in sport and rehabilitation. Sensors 22(8). https://doi.org/10.3390/s22083006

Jensen U (2016) Design considerations and application examples for embedded classification systems. Dissertation, FAU University Press, Erlangen, Germany, parallel erschienen als Druckausgabe bei FAU University Press, ISBN: 978-3-944057-86-6

Jin X, Li L, Dang F, et al (2022) A survey on edge computing for wearable technology. Digital Signal Processing 125:103146. https://doi.org/https://doi.org/10.1016/j.dsp.2021.103146

Jácome C, Ravn J, Holsbø E, et al (2019) Convolutional neural network for breathing phase detection in lung sounds. Sensors 19(8). https://doi.org/10.3390/s19081798

Kos A, Milutinović V, Umek A (2019) Challenges in wireless communication for connected sensors and wearable devices used in sport biofeedback applications. Future Generation Computer Systems 92:582–592. https://doi.org/https://doi.org/10.1016/j.future.2018.03.032

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Advances in neural information processing systems 30

McClure K, Erdreich B, Bates JHT, et al (2020) Classification and detection of breathing patterns with wearable sensors and deep learning. Sensors 20(22). https://doi.org/10.3390/s20226481

Mencarini E, Rapp A, Tirabeni L, et al (2019) Designing wearable systems for sports: A review of trends and opportunities in human–computer interaction. IEEE Transactions on Human-Machine Systems 49(4):314–325. https://doi.org/10.1109/THMS.2019.2919702

Murphy MJ, Pokhrel D (2009) Optimization of an adaptive neural network to predict breathing. Medical Physics 36(1):40–47. https://doi.org/https://doi.org/10.1118/1.3026608

Novacheck TF (1998) The biomechanics of running. Gait & Posture 7(1):77–95. https://doi.org/10.1016/S0966-6362(97)00038-6

Pain MTG, Hibbs A (2007) Sprint starts and the minimum auditory reaction time. Journal of Sports Sciences 25(1):79–86. https://doi.org/10.1080/02640410600718004, pMID: 17127583

Paul HA (2017) Biofeedback: A practitioner's guide, fourth edition, edited by m. schwartz & f. andrasik. Child & Family Behavior Therapy 39(2):161–170. https://doi.org/10.1080/07317107.2017.1307683

Postma DB, van Delden RW, Koekoek JH, et al (2022) A design space of sports interaction technology. Foundations and Trends® in Human–Computer Interaction 15(2-3):132–316. https://doi.org/10.1561/1100000087

Ramrath L, Schlaefer A, Ernst F, et al (2007) Prediction of respiratory motion with a multi-frequency based extended kalman filter. In: Proceedings of the 21st international conference and exhibition on computer assisted radiology and surgery (CARS'07), pp 56–58

Rosenblum M, Pikovsky A, Kühn AA, et al (2021) Real-time estimation of phase and amplitude with application to neural data. Scientific Reports 11(1):18037. https://doi.org/10.1038/s41598-021-97560-5

Schranz C, Halmich C, Mayr S, et al (2024) Surrogate modelling of heartbeat events for improved j-peak detection in BCG using deep learning. Frontiers in Network Physiology 4. https://doi.org/10.3389/fnetp.2024.1425871, publisher: Frontiers

Semenkov I, Fedosov N, Makarov I, et al (2023) Real-time low latency estimation of brain rhythms with deep neural networks. Journal of neural engineering 20(5). https://doi.org/10.1088/1741-2552/acf7f3

Singh R, Gill SS (2023) Edge ai: A survey. Internet of Things and Cyber-Physical Systems 3:71–92. https://doi.org/https://doi.org/10.1016/j.iotcps.2023.02.004

Smetanin N, Belinskaya A, Lebedev M, et al (2020) Digital filters for low-latency quantification of brain rhythms in real time. Journal of Neural Engineering 17(4):046022. https://doi.org/10.1088/1741-2552/ab890f

Tan M, Le Q (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol 97. PMLR, pp 6105–6114

Technologies I (2024) ModusToolbox Machine Learning User Guide. URL https://www.infineon.com, version 09.00, Accessed: 2024-01-26

Teerapittayanon S, McDanel B, Kung HT (2016) Branchynet: Fast inference via early exiting from deep neural networks. In: 2016 23rd international conference on pattern recognition (ICPR), IEEE, pp 2464–2469

Umek A, Kos A (2016) The role of high performance computing and communication for real-time biofeedback in sport. Mathematical Problems in Engineering 2016(1):4829452. https://doi.org/https://doi.org/10.1155/2016/4829452

Valsted FM, Nielsen CVH, Jensen JQ, et al (2017) Strive: exploring assistive haptic feedback on the run. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction. Association for Computing Machinery, New York, NY, USA, OzCHI '17, p 275–284, https://doi.org/10.1145/3152771.3152801

Vasilyev V, Borisov V, Syskov A (2019) Biofeedback methodology: a narrative review. In: 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), pp 0011–0016, https://doi.org/10.1109/SIBIRCON48586.2019.8958019

Virtanen P, Gommers R, Oliphant TE, et al (2020) Scipy 1.0: fundamental algorithms for scientific computing in python. Nature Methods 17:261–272

Vlemincx E, Abelson JL, Lehrer PM, et al (2013a) Respiratory variability and sighing: A psychophysiological reset model. Biological Psychology 93(1):24–32. https://doi.

org/https://doi.org/10.1016/j.biopsycho.2012.12.001

Vlemincx E, Vigo D, Vansteenwegen D, et al (2013b) Do not worry, be mindful: Effects of induced worry and mindfulness on respiratory variability in a nonanxious population. International Journal of Psychophysiology 87(2):147–151. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2012.12.002