Using LSTM Networks and Future Gradient Values to Forecast Heart Rate in Biking

Henry Gilbert, Jules White, Quchen Fu, Douglas C. Schmidt
[1] *Vanderbilt University*
henry.gilbert@Vanderbilt.edu, jules.white@Vanderbilt.edu, quchen.fu@Vanderbilt.edu, douglas.c.schmidt@vanderbilt.edu

Keywords: Deep LSTM, Deep Neual Network, Heart Rate, Forecasting, Biking

Abstract: Heart Rate prediction in cycling potentially allows for more effective and optimized training for a given individual. Utilizing a combination of feature engineering and hybrid Long Short-Term Memory (LSTM) models, this paper provides two research contributions. First, it provides an LSTM model architecture that accurately forecasts the heart rate of a bike rider up to 10 minutes into the future when given the future gradient values of the course. Second, it presents a novel model success metric optimized for deriving a model's accuracy to predict heart rate while an athlete is zone training. These contributions provide the foundations for other applications, such as optimized zone training and offline reinforcement models to learn fatigue embeddings.

1 INTRODUCTION

Monitoring intensity during exercise and training is essential to optimize performance (Sylta et al., 2014). Insufficient intensity during training yields slower or negligible performance progression. Excessive intensity, however, yields over-training and the potential for injury or performance degradation (Collinson et al., 2001).

There are several metrics for measuring exercise intensity, including heart rate, V02 max, and power output (Dooley et al., 2017). Heart rate can be consistently measured across all types of athletics, whereas power output is exclusive to cycling. Likewise, heart rate can be used as a reliable indicator of exercise intensity (Jeukendrup and Diemen, 1998). Our intensity prediction efforts therefore focus on future heart rate.

Yet, predicting heart rate on a given cycling course is a nuanced and difficult problem. In particular, there are various external factors that can not be accounted for, even given the geographical data for a course. For example, some materials require greater effort to peddle across based on the material's composition. Moreover, even excluding external course factors, there is a need to account for cardiac drift, which is a continual rise or decline of heart rate after exercise due the body's internal temperature. (Dawson, 2005). In addition, there is inconsistency amongst humans, *e.g.*, any data derived for training must account for the inevitable invariance caused by a person's inability to keep perfect pace or other intangible internal factors.

Research Question: Can deep Long Short-Term Memory (LSTM) models be used to forecast biker heart rates when given the future gradient values of the course? The ability for a model to accurately forecast heart rate in the future can potentially help improve the way athletes approach zone training. Athletes today typically train reactively, *i.e.* if during exercise their heart rate drops too low, they increase intensity. If their heart rate becomes too high, they decrease intensity (NEUFELD et al., 2019). This approach can be a sub-optimal and yield a constantly oscillating heart rate and a subsequently greater amount of time spent outside the correct zone for training.

In contrast, if athletes can accurately predict their heart rate minutes into the future, they can proactively adjust their intensity to stay within the desired zone. For example, if a model forecasts an athlete's heart rate will spike out of a given zone due to an upcoming hill in two minutes, the athlete can proactively decrease their effort to lower their heart rate to prepare for the increased effort of the hill. By limiting heart rate oscillation, the time in the correct zone can increase, thereby optimizing training performance.

Moreover, future heart rate prediction can be applied to an offline learning model when given enough contextual data. For example, a model can learn the embedding of intensity for a given athlete and apply that learning to optimally indicate when the athlete should speed up or slow down to ensure their heart rate remains in a given zone. This is a novel approach that fundamentally contrasts with a coach reactively instructing an athlete. Therefore, an independent model can optimally and proactively dictate an athlete's intensity. This information could be communicated to an athlete in the form of a [1,10] intensity scale, with 1 being minimum effort and 10 being maximum effort, thereby potentially maximizing training for an athlete's specific intensity embeddings.

Key Contribution: A promising deep LSTM Model for predicting heart rate Up to 600 seconds in the future. This paper presents an architecture for an LSTM model (Hochreiter and Schmidhuber, 1997) that can predict heart rate up to 600 seconds (*i.e.*, 10 minutes) in the future. This LSTM model achieves this predictability via a mix of past historical heart rate and velocity data with the future gradient values of the course n minutes into the future.

The remainder of this paper is organized as follows: Section 2 summarizes related work and explores its relationship with the research described in this paper; Section 3 illustrates the models underlying architecture and the reasoning behind it; Section 4 explains the model's performance in relation to the experimental data and analyzes the experiment results; and Section 5 presents concluding remarks.

2 Related Work

Prior work exploring heart rate prediction using machine learning can be categorized as focusing on (1) predicting heart rate given noisy data and (2) forecasting heart rate in the context of sports. This section summarizes this related work and explores its relationship with the research described in this paper.

An important problem in the analysis of heart rate is the quality of the data being collected. Unfortunately, due to external factors (such as sweat and excessive motion during exercise) common sensors (such as PPG and ECG) fail to produce accurate readings. Recent work has applied machine learning to forecast heart rate more accurately, regardless of the noise within the input data.

For example, (Yun et al., 2018) focused on forecasting heart rate variability via a Hidden Markov model. Likewise, (Fedorin et al., 2021) compared sequence-to-sequence models and deep LSTM models with integrated CNN layers for more accurate heart rate forecasting and concluded that LSTM model performed marginally better than the sequence-to-sequence model. In contrast, our work focuses on a related—but different—application, ɪ.e., ingesting and forecasting using noisy data to predict heart rate in the context of biking, rather than subjects at rest.

A recent paper (Staffini et al., 2022) uses the past n minutes of heart rate data to predict heart rate at $n+1$. Using only an auto-regressive model, this paper achieved an mean absolute error (MAE) of 3 bmp. However, once again, that paper's data set consisted entirely of subjects at rest. Thus, their model allowed less variance in the external environment when compared to heart rate forecasting in an exercise setting, such as our focus on biking in this paper.

Yet another recent paper (Ni et al., 2019) combined LSTM with sequence-to-sequence and encoding layers to create a workout forecasting model, a short-term heart rate forecaster, a linear projection embedding module, and a workout recommender system. Given contextual information about the user's historical performance and sport type, they constructed a model to forecast the user's heart rate for a given workout route at minute intervals. The short-term forecasting model used a custom encoder-decoder with temporal attention to forecast heart rate 10's ahead with a resulting root means square error (RMSE) of 7.025 heart beats. In contrast, our work focuses on heart rate in the context of biking and specifically applying future gradient values of a given course for predictions up to 600 seconds ahead.

Finally, one of our earlier papers (Qiu et al., 2021) is based on the same data set as this paper and also forecasts heart rate at a one second interval using several models, including Random Forest, a Feed-Forward Neural Network (FFNN), a Recurrent Neural Network (RNN), and LSTM. That paper found the LSTM model exhibited the top forecasting performance out of the aforementioned models. While that paper shared a similar research question and the same data set as our current paper, the papers differ in two ways: (1) our current paper introduces a novel forecasting performance metric to relate predictions to their heart rate zone and (2) it also applies future gradient values to increase the LSTM model's predictive power and ability to generalize.

3 Heart Rate Prediction Model Architecture

This section describes the LSTM architecture that we analyzed in our experiments to predict future heart rate values given the future gradient values. Figure 1 shows the overall architecture for the LSTM model that produced the best results, which consisted of 5 layers of 1000 neurons each. We experimented with a variety of layers and neurons. Specifically, we experimented with models ranging from 2 layers of 50 neurons to models with 10 layers of 5,000 neurons. Two additional dense layers were appended to the end

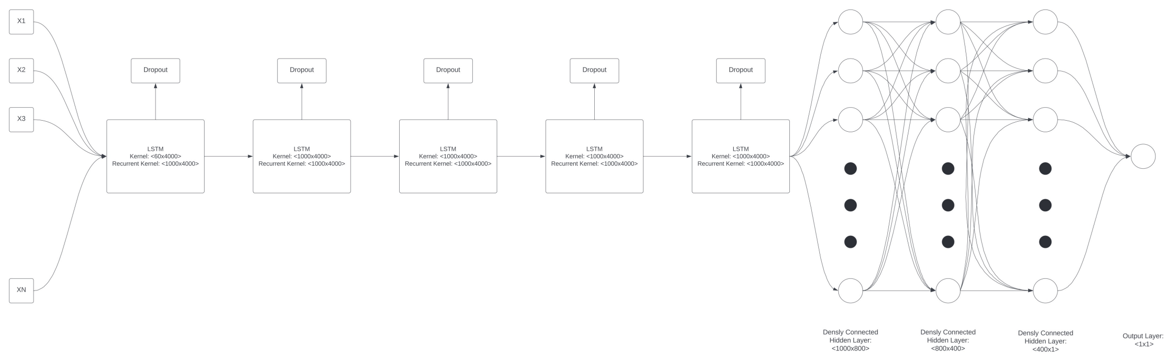

Figure 1: Model Architecture

of the model to learn the output embedding from the LSTM layers and apply it to forecast the heart rate.

Our experiments indicated that smaller models were unable to capture the complexity of the relationships present in the data. Such models would often end up converging to the average heart rate of the training set, rather than a mapped relationship between gradient and heart rate. Conversely, models larger than 5 layers of 1,000 neurons often had diminishing returns and struggled with over fitting the training data. Thus, they were not worth the exponential increase in compute time.

For heart rate prediction, our LSTM model used three input vectors: past heart rate, past cadence, and future gradient. All data was normalized between [0,1]. Each of these vectors were bucketed and averaged by chunks of five to reduce the input space and limit the noise found in such granular data.

Chunk size was determined via grid search experimentation, where we experimented with values in the range of [1,2,5,10,50,100]. A chunk size of five conclusively performed the best. In particular, this chunk size optimally balanced between retained information density and noise reduction through smoothing.

These three vectors were then interwoven into each other such that given vector heart rate = $[h_0, h_1, h_2]$, vector cadence = $[c_0, c_1, c_2]$ and vector gradient = $[g_3, g_4, g_5]$. The resulting vector was then $[h_0, c_0, g_3, h_1, c_1, g_4 h_2, c_2, g_5]$. The heart rate and cadence are historical values, while the gradient is the future gradient of the route.

This interwoven list was subsequently segmented into chunks of six, such that the model took a depth of six steps for every feature vector x ∈ X. Given historical data h_0, ... ,h_n, c_0, ... ,c_n and g_0, ... ,g_n, the model predicted the heart rate at time step 2n. Specifically, for a prediction 120 seconds into the future, the input was reformatted into chunks of the previous 120's of heart rate and cadence, along with the future 120's of the gradient. Given 120's chunk of historical data, $[t_0, t_{120}]$ was used to make a singular prediction, 120's into the future at time step t_{240}. Thus, given heart rate = $[h_0, ..., h_{120}]$, cadence = $[c_0, ..., c_{120}]$ and gradient = $[g_{120}, ..., g_{240}]$, the model predicted the heart rate of the athlete at time step t_{240}.

Our approach interwove past heart rate, past cadence, and future gradient together to enable the segmentation of time steps when the data was fed to the model. If heart rate, cadence, and gradient were merely appended to the end of each other that data could not be segmented into time steps because the first step would consist entirely of heart rate data, the last step would consistent entirely of gradient values, and the rest would be a mix of all three. To achieve consistent and homogeneous time steps the three data vectors must be interleaved.

Our LSTM model's input layer took batch sizes of one, where each time-step had a feature vector of size of size 10. There were $model - look - forward - amount/10$ time-steps for each input into the model. This part of the algorithm yielded a many-to-one LSTM input architecture.

This input was then fed into four subsequent and identical LSTM layers, each with 1,000 neurons. Every one of these hidden LSTM layers had a 0.2 integrated drop layer. This portion of the algorithm yielded a trainable 800 neurons, per layer, per pass.

The output of the deep LSTM model was then fed into a densely connected layer of 800 neurons. The subsequent output from this layer was next fed into two further densely connected hidden layers with 300 neurons each. Finally, the output was condensed into a final single neuron with a linear activation applied to forecast the heart rate.

Our LSTM model utilized a mean absolute error loss function. Specifically, the absolute value of the forecasting errors are summed together and used to adjust the networks weights, as shown in Equation 3.

$$\sum_{i=1}^{D} |x_i - y_i|$$

As described previously, the model also applied a custom metric to determine model forecasting success in relation to heart rate zones. Specifically, to calculate the error between actual heart rate and predicted heart rate, this metric created a relative error to the size of the heart zone, which distinguished between upper and lower bounds of a given zone and related the distinction to the actual size of that zone.

For example, given a heart rate zone between 200 and 180 beats per minute the size of this heart rate zone is 20 beats. Therefore, if the actual heart rate is 185 and the predicted heart rate is 190 the actual error is 5 beats. Given this zone size 20, the adjusted error is 5/20 or 0.25.

In contrast, for a heart zone between 200–190 with a zone size of 10 the actual error is also 5 beats if the actual heart rate is 195 and the predicted heart rate is 200. In this case, however, the adjusted error becomes 5/10 or 0.5. Since the zone is smaller, the relative difference between actual and predicted heart rate carries a greater impact.

The difference in importance using the concrete error cannot be captured in this scenario. However, the relative adjusted error *does* capture this discrepancy in importance. This distinction is shown by the concrete error being 5 beats for both examples, but a higher adjusted error of 0.5, rather than 0.25. This difference compensates for the narrower margins of the smaller heart rate zone.

4 Experiment Methods and Results

This section gives an overview of our experiment data and training method, analyzes the results from our experiments, and explains threats to validity.

4.1 Overview of the Experiment

The data set used to train LSTM model comprised the grade of the course, speed, heart rate, elevation, distance and cadence, all measured at one second intervals. There were 110 overall rides, resulting in a data set of 270,000 examples. We used both a validation set of 30k examples and the same validation set described in (Qiu et al., 2021) to directly contrast model performance more accurately.

Our LSTM model was trained on a machine with 48GB of VRAM and 256GB RAM. This model utilized early stopping with a patience of five and a minimum value loss variance of one. The model ran for an average of eight epochs at a training time of four minutes each, resulting in an average training time of 32 minutes.

4.2 Analysis of Results

A concrete example of the model's performance on a single ride is shown in Figure 2. Exact model performance for each ride in the validation set is shown in Table 1.

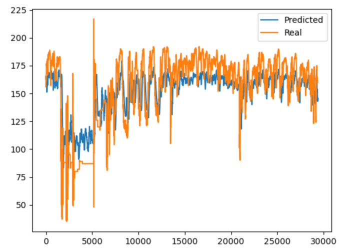

Figure 2: Model Results for 120's prediction across validation set of 30,000 data points

Validation Ride	MAE	RZE
Run 0	13.82	0.47
Run 1	21.61	1.07
Run 2	23.23	1.04
Run 3	22.99	0.86
Run 4	32.34	1.52
Run 5	19.17	0.83
Run 6	14.84	0.58
Run 7	23.29	0.71
Run 8	22.67	1.02
Run 9	32.54	0.98

Table 1: Model Performance on Validation Rides

The corresponding run graph results are shown in Figures 3, 4, 5, 6, 7, 8, 9, 10, 11 12. Given the validation set of 30k examples, the model achieves a mean

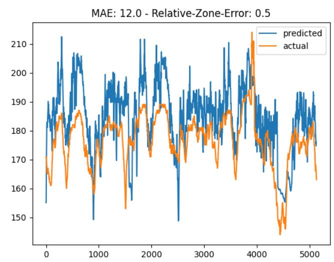

Figure 3: Model Results for 120's Prediction on Validation Run 0

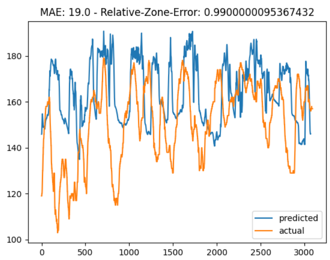

Figure 4: Model Results for 120's Prediction on Validation Run 1

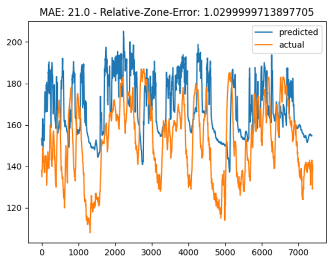

Figure 5: Model Results for 120's Prediction on Validation Run 2

squared loss of 25.63, though there are outliers around step 5,000 in the validation set. Regardless of the inherent noise in the data, the model continues to fit the general trend and not diverge to only predicting the most recent heart rate. These results showcase the model's ability to learn the relationship between current heart rate and future gradient values.

This same model can then be retrained to predict 10 minutes into the future. Even predicting so far into the future, the model still generalizes to the correct trend, as shown in Figure 13.

The 10 minute model's predictions fail to continually match the trend of the users heart rate across the entirety of the 2,000 time steps. However, this result shows the model's ability to learn the strong relationship between past heart rate and cadence values

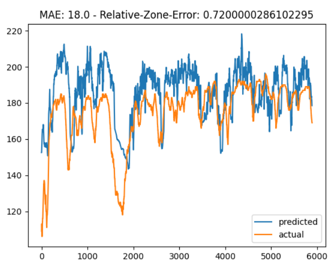

Figure 6: Model Results for 120's Prediction on Validation Run 3

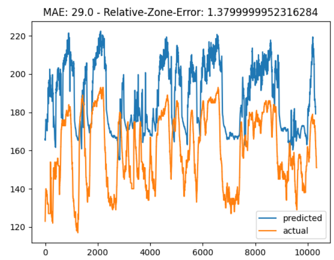

Figure 7: Model Results for 120's Prediction on Validation Run 4

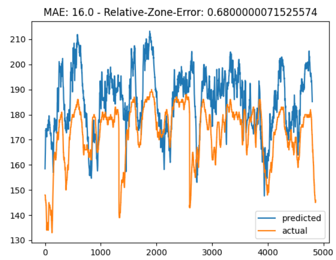

Figure 8: Model Results for 120's Prediction on Validation Run 5

to future gradient values. Moreover, the divergence in accuracy between time steps 750 and 1,250 can be explained by an analysis of the user's cadence during the same time period. As shown in Figure 14, the cadence drops to 0, suggesting the athlete stopped during the user's ride.

Naturally, the model can not account for an unexpected departure of the course and the subsequent cardiac drift that ensues. However, the model was quickly able to refit the trend, as shown in Figure 13. Moreover, Figure 15 depicts forecasting heart rate at 120's intervals in relation to the associated heart rate zones for the given athlete.

This model can therefore make valuable and actionable predictions if it is able to fit the trend of heart rate zones, which is not necessarily a direct trend

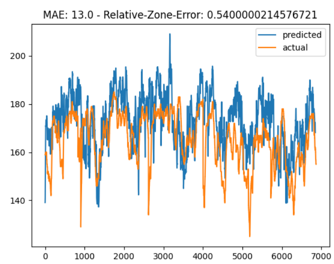

Figure 9: Model Results for 120's Prediction on Validation Run 6

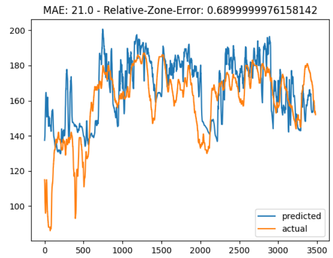

Figure 10: Model Results for 120's Prediction on Validation Run 7

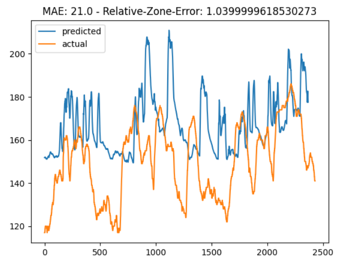

Figure 11: Model Results for 120's Prediction on Validation Run 8

with the heart rate. It is also essential to consider the model's confidence bounds, as shown in Figure 16.

Allowing for a certain margin of uncertainty in the model's predictions yields a better generalization towards predicting zone relations. Specifically, this relationship can be quantified to zone trends by taking the absolute difference in a model's prediction and the ground truth and then scaling the difference to the size of an individuals heart rate zone. For reference, the model seen in 15 has an average relative zone error of 0.414, so it is accurate to within 60 percent of a given heart zone for any x ∈ X. Figure 13 and Figure 17.

An example of the importance for generalized zone forecasting instead of granularity heart rate is shown in 13 It initially appeared the 10 minute model was unable to fit the validation set as well as as the

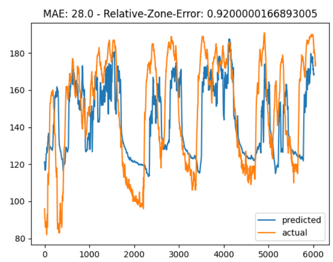

Figure 12: Model Results for 120's Prediction on Validation Run 9

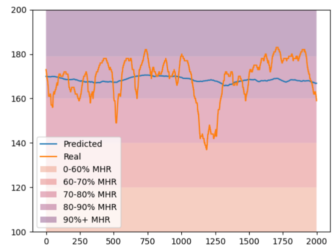

Figure 13: Model Results for 600's Predictions

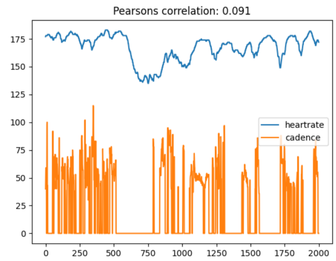

Figure 14: Cadence in Relation to Heart Rate

two minute model. However, the 10 minute model sacrificed granular tracking for a more robust relationship with the given zone for a heart rate. This trade off allowed the model to achieve a lower relative zone error score of 0.37.

4.3 Threats to Validity

The results reported above show that our LSTM model's utility for optimized heart rate zone training is promising and potentially actionable. However, the current data set consists of only one athlete cycling on a multitude of different routes and has not yet been tested on other individuals cycling on more diverse routes and in other sport types, such as running. Intuitively, this model should generalize to a greater population of subjects and contexts as the model's inputs are agnostic of the route, person or sport type. It also

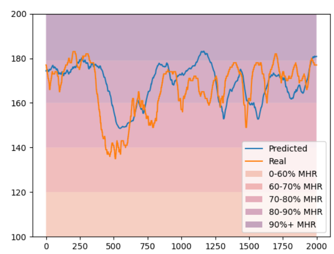

Figure 15: 120's Prediction for Single Ride in Reference to Training Zones

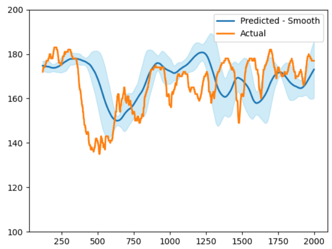

Figure 16: 120's prediction in reference to it's confidence intervals

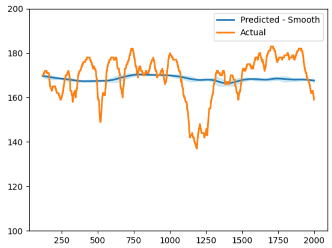

Figure 17: 600's prediction in reference to it's confidence intervals

remains an open research question if the model can generalize across an entire population of athletes or if each individual athlete will need a model fine-tuned to their internal exertion embeddings.

5 Concluding Remarks

Few papers have been published that provide deep learning model architectures for forecasting heart rate more than 60 seconds into the future. As discussed in Section 2, even fewer proposed models attempt to forecast heart rate in an athletic context.

This paper showed that a hybrid LSTM model in conjunction with supplying future gradient values for a given course allows prediction of actionable heart rate forecasts up to 10 minutes into the future. Specifically, by feeding a deep LSTM model into a densely connected deep neural network, we are able to accurately forecast the heart rate of an athlete biking in relation to the gradient values of a given course. These results suggest other potential applications for optimized zone training and offline learning for athlete intensity management through exertion embeddings.

Our research reported in this paper yield the following lessons regarding forecasting heart rate in an athletic context:

1. **LSTM architecture for heart rate forecasting**. As shown in this paper, the LSTM model architecture is an effective choice for forecasting future heart values.
2. **Increased forecasting ability with future gradient values**. Our LSTM model effectively predicted the heart rate of an athlete up to 10 minutes into the future when using the future gradient values of the course. This result demonstrated the strong predictive relationship between the gradient of a course and heart rate.
3. **Adjusted model success metrics**. We learned that adjusting the model prediction error in relation to the size of a heart rate zone is a more descriptive metric to define model success.

Our future work will continue evolving this research by performing larger experiments with a more diverse subject group to gain a greater breadth of data. Specifically, we will increase our sample size by studying 20 different athletes of an even distribution of gender, as well as have each athlete cycle 10 different routes and run the same 10 routes. We will then use this data to retrain a larger version of the current LSTM model and test whether this model can effectively learn the relationship between future gradient and current heart rate, irrespective of an athlete's gender, fitness level, or sport.

All of the models and data that we presented in this paper can be downloaded in open-source form at the following GitHub repository URL: `github.com/henrygilbert22/RL-Human-Performance`. Please consult the repo's ReadMe file for in-depth instructions on replicating this paper's results and a more extensive description of the data used in the experiments.

REFERENCES

Collinson, P. O., Boa, F. G., and Gaze, D. C. (2001). Measurement of cardiac troponins. volume 38, pages 423 – 449.

Dawson, E. A. (2005). Cardiac drift during prolonged exercise with echocardiographic evidence of reduced diastolic function of the heart. volume 94, page 3.

Dooley, E. E., Golaszewski, N. M., and Bartholomew, J. B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR Mhealth Uhealth*, 5(3):e34.

Fedorin, I., Slyusarenko, K., Pohribnyi, V., Yoon, J., Park, G., and Kim, H. (2021). Heart rate trend forecasting during high-intensity interval training using consumer wearable devices. In *Proceedings of the 27th Annual International Conference on Mobile Computing and*

Networking, MobiCom '21, page 855–857, New York, NY, USA. Association for Computing Machinery.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Jeukendrup, A. and Diemen, A. V. (1998). Heart rate monitoring during training and competition in cyclists. volume 16, pages 91–99.

NEUFELD, E., WADOWSKI, J., BOLAND, D., Dolezal, B., and Cooper, C. (2019). Heart rate acquisition and threshold-based training increases oxygen uptake at metabolic threshold in triathletes: A pilot study. *International Journal of Exercise Science*, 12:144–154.

Ni, J., Muhlstein, L., and McAuley, J. (2019). Modeling heart rate and activity data for personalized fitness recommendation. In *The World Wide Web Conference*, WWW '19, page 1343–1353, New York, NY, USA. Association for Computing Machinery.

Qiu, X., White, J., and Schmidt, D. C. (2021). Analyzing machine learning models that provide personalized heart rate forecasting for elite cyclists.

Staffini, A., Svensson, T., Chung, U.-i., and Svensson, A. K. (2022). Heart rate modeling and prediction using autoregressive models and deep learning. volume 22.

Sylta, , Tønnessen, E., and Seiler, S. (2014). From heart-rate data to training quantification: A comparison of 3 methods of training-intensity analysis. *International journal of sports physiology and performance*, 9:100–7.

Yun, S., Son, C.-S., Lee, S.-H., and Kang, W.-S. (2018). Forecasting of heart rate variability using wrist-worn heart rate monitor based on hidden markov model. In *2018 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–2.