A Study of Machine Learning Models for Personalized Heart Rate Forecasting in Mountain Biking

Xiaoxing Qiu[1], Jules White[1] and Douglas C. Schmidt[1]
[1]*Department of Computer Science, Vanderbilt University, Nashville, TN 37212 USA*
{*xiaoxing.qiu, jules.white, d.schmidt*}*@vanderbilt.edu*

Keywords: Heart rate forecasting, personalization, cycling course, deep learning

Abstract: Performance forecasting in cycling is most effective when it is personalized and course-specific to account for the influence of individual and terrain factors. This paper empirically assesses various personalized and course-specific performance forecasting models based on four machine learning models, including random forest, feed forward neural networks (FFNNs), recurrent neural networks (RNNs), and long short term memory (LSTM). The mean square error (MSE) is selected as the metric for model comparison. The results of our experiments show that despite the severely overfitted random forest models the LSTM models have the lowest MSE in the heart rate forecasting on our test dataset.

1 INTRODUCTION

1.1 Background

Monitoring exercise intensity during cycling is important. While insufficient exercise may slow improvement in an athlete's performance, excess exertion can result in over-training or even muscle damage. Overtraining in cycling can result in a debilitating syndrome that degrades the performance of cyclists for several months and may ultimately result in failure to meet competition goals (Gleeson, 2002). Moreover, over-trained athletes—especially those involved in endurance sports like cycling—are susceptible to infections and require significantly longer time for recovery than non-athletes (Gleeson, 2002; Nieman, 1994).

Several metrics are available to estimate exercise intensity, including heart rate, power output, and $\dot{V}O_2max$. According to Jeukendrup et al's study (Jeukendrup and Diemen, 1998), heart rate is a reliable indicator of cycling exercise intensity since it is largely independent of a specific course. However, heart rate can be affected by terrain factors, such as grade of the hill, which varies on different riding courses.

Predicting heart rate at different points in time on a cycling course is hard since heart rate is affected by course-specific features, as well as other personalized physiological factors. Research has shown that a cyclist's heart rate drifts upwards after exercising for 20 to 60 minutes despite unchanged work loads (Bian et al., 2019; Le et al., 2009), which is a condition called "cardiac drift."

Cardiac drift is associated with an increase of core body temperature during exercise (Dawson et al., 2005; Collinson et al., 2001), which may cause athletes to lower their speed to maintain their heart rate. On the other hand, higher speed can yield a higher heart rate in a given condition (Le et al., 2009). Moreover, heart rate responses vary with a cyclist's individual factors, such as gender and age (Le et al., 2009), as well as their cadence on different parts of a given course.

It is hard to predict heart rate on a specific point in a course at a target speed. Training plans therefore focus on setting target heart rate or power goals (e.g., power applied to the pedals by the rider) and riding as fast as possible without exceeding those targets. Ideally, training plans could be developed where heart rate at a given speed and point in a course is predicted and riders are given a set of target speeds to ride instead. Achieving this ideal, however, requires building heart rate forecasting models that are personalized and course-specific, which yields the following benefits:

- **Achieve speed goals without over-training.** A personalized performance forecasting model for specific courses is vital for cyclists to establish achievable speed goals at different courses in advance and maximize training effects without over-training.

- **Improve course-specific estimation accuracy.** Riders can accurately estimate how much faster they could ride in different sections of the course, yet still remain within heart rate targets. For example, a rider who has only ridden a course at 70-80% of max heart rate could estimate how fast they would ride at a race intensity of up to 90% of max heart rate. This estimate could give them a benchmark of what they could currently achieve without actually riding the course and potentially over-exerting themselves.

A biker may need to climb up on a steep uphill trail on uneven terrain with 6-8in rocks, which requires significantly slower speeds and greater intensity than a similar climb on smooth ground. Similarly, a high-speed downhill ride on smooth terrain will place less stress on a rider's core muscle groups compared to the same downhill speed across tree roots. The individual terrain features, turns, gradients, and other aspects of a course significantly impact a rider's speed and intensity, but current models are mainly course-independent.

To maximize their efficacy, personalized and course-specific heart rate models should be easily trainable from limited data—ideally a single test ride of a course by an athlete. If an athlete rides a course multiple times at multiple target heart rates, they are less likely to need a predictive model since they already know the course well. Therefore, new courses (i.e., where the athlete has limited knowledge) are those where personalized and course-specific predictive models are most valuable. Ideally, a cyclist should be able to ride a course once and then predict how different target speeds would impact heart rate at different points on the course.

1.2 Research Question: Which machine learning approaches perform best for personalized course-specific heart rate prediction from a single ride?

This paper presents our research on comparing personalized multivariate models to forecast the heart rate of a cyclist on a specific course using data from a single ride. These models consider course-specific factors at each part of the course (such as the grade of road and the altitude), as well as current rider details (such as the cadence), and then forecast the heart rate of the athlete based on them. We compare the results of heart rate and speed forecast by the following machine learning models from a single ride and report which models perform best:

- The first type of model experimented with is *random forest*, which is a traditional machine learning model. Forecasting results show that random forest models have severe overfitting and therefore cannot be utilized in heart rate forecasting from a single ride.
- To mitigate the random forest overfitting problem, we implement *feed forward neural networks* (FFNNs). FFNNs do not exhibit overfitting problems on the dataset, though their forecasting accuracy is lower. In particular, FFNNs do not consider historical information when forecasting heart rate.
- To address the limitations with FFNNs, we also implemented two other types of neural networks: *simple recurrent neural networks* (RNNs) and *long short term memory* (LSTM) networks.

We empirically compare all these models via experiments in our testbed. The results of these experiments indicate that the LSTM models have the lowest mean square error among machine learning models despite severely overfitted random forest models and produce predictions that closely match real-world heart rate sequences.

1.3 Paper Organization

The remainder of this paper is organized as follows: Section 2 describes prior work on heart rate forecasting; Section 3 reviews the data processing methods and all the models applied in this paper, including basic concepts of each model and their pros and cons; Section 4 describes how we built random forest, feed forward neural network (FFNN), recursive neural network (RNN), and long short term memory (LSTM) models to forecast the heart rate of athletes on a specific route and then compares the best models among these four types with two course-independent models (FitRec (Ni et al., 2019) and Minmin's LSTM-based model (Luo and Wu, 2020)); Section 5 summarizes the results from our experiments in Section 4; and Section 6 presents concluding remarks and outlines future work.

2 RELATED WORK

Researchers have built various models to predict the performance of elite cyclists. Le et al. (Le et al., 2009) proposed a mathematical model to evaluate athletes' heart rate response under moderate exercise intensity based on physical and physiological principles. Lucia et al. (Lucía et al., 2001) analyzed the preferred cadence of elite cyclists and found that on flat stages they tend to adopt higher cadences (around 90 rpm) while on mountain ascents cadences are around 70 rpm. However, their models focus on laboratory conditions. Course-specific factors, such as the slope of the road in real courses, are not consider, though these factors significantly influence a cyclist's heart rate response.

Due to the advent of wearable devices, large amounts of data can be collected and processed via mobile devices, which offers an opportunity to build personalized performance forecasting models. Mohammadzadeh et al. (Mohammadzadeh et al., 2018) applied a support vector machine predictor to predict the breathing rate based on the 3-D accelerations, heart rate, body temperature, electrodermal activity, humidity etc. in a controlled lab environment. Ming et al. (Ming and Jun, 2008; Xiao et al., 2010) used an FFNN to investigate the relationship between heart rate and physical activity in daily life with the help of a wearable physical activity recorder that monitors the 3-D accelerations of the body.

RNNs exhibit sequential correlation and can seamlessly model problems with multiple inputs. These models are therefore widely used in natural language processing and time series prediction (Cho et al., 2014). In athletic performance forecasting, RNN based models can take personalized factors (such as blood pressure and running speed during exercise) to make a heart rate predictions. Ni et al. (Ni et al., 2019) proposed an LSTM-based model to learn a user's heart rate profile during exercise and offer workout route recommendation and short term heart rate prediction. Luo et al. (Luo and Wu, 2020) also proposed an LSTM-based model to predict heat rate based on heart rate signal, gender, age, accelerations and mental state. Bian et al. (Bian et al., 2019) tracked facial key points from each frame of facial videos to estimate heart rate.

In summary, although there are other performance forecasting models, most studies focus on the heart rate profile collected either during daily activities or under laboratory conditions. There are few models that are course-specific and personalized to forecast a cyclist's heart rate and speed. However, such models can be beneficial to both cyclists and coaches.

For example, a cyclist needs a model to establish various speed goals at different parts of a course before a competition or predict their heart rate based on given speed goals. A coach can use such a model to evaluate the heart rate of a given athlete on a given course at a given speed to ensure the exercise intensity and avoid over-training. Likewise, when given specific heart rate goals, these models can be used to predict how fast a cyclist can/should ride at different parts of a course.

3 METHODOLOGY

To address the limited understanding of personalized and course-specific heart rate forecasting from a single ride, we evaluated the performance of prior course/cyclist-independent models, course/cyclist-specific traditional machine learning models, and course/cyclist-specific neural networks on forecasting rider heart rate on a single ride of a course. These comparisons allowed us to investigate a number of important research questions and collect important lessons learned to guide future research, as discussed in this section.

3.1 Key Research Questions

The key research questions we investigated in our study include the following:

- Are current models that do not consider course-specific features as good as models that consider specific features, such as location?
- Do traditional machine learning or neural networks perform better on course-specific heart rate forecasting?
- For traditional machine learning models, which features are most salient for learning?

3.2 Experimentation Approach

We began our study by surveying prior work on heart rate forecasting. We then selected and applied both traditional machine learning models and neural networks to a cycling dataset that we collected. Important cyclist and course-specific factors must be considered for cycling performance forecasting. It is natural to consider multivariate models for performance forecasting, e.g., random forest and neural networks are popular machine learning algorithms because they work for both regression and classification and can handle multiple inputs. The flow chart in Figure 1 outlines key procedures associated with heart rate forecasting.

According to Leijnen et al. (Leijnen and van Veen, 2020), there are 13 major neural network architectures used by researchers. Among all these neural networks, three of them are widely used in performance forecasting, including feed forward neural networks (FFNN) (Ming and Jun, 2008; Xiao et al., 2010), basic recurrent neural networks (RNN) (Chowdhury

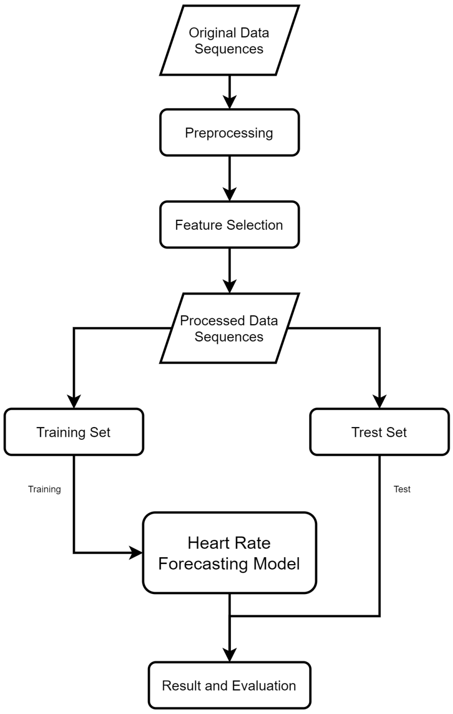


Figure 1: Heart Rate Forecasting Procedures

et al., 2019), and long short term memory (LSTM) (Bian et al., 2019; Ni et al., 2019) models. We therefore selected these three neural networks and random forest and built a personalized model to predict the performance of a cyclist on a specific routes. LSTM is a special type of RNN, so to distinguish these two models we call basic RNN models "simple RNN" models in this paper.

In the performance forecasting problem, an athlete's performance is affected by numerous features, such as temperature and the grade of course. Not all these factors, however, should be provided as input to a traditional machine learning model. For example, random forest models are sensitive to data variation, so small differences in the dataset can cause a large variance in the prediction.

Pruning irrelevant features to an athlete's performance can reduce model overfitting and improve forecasting accuracy. Neural networks have built-in mechanisms to mitigate overfitting and perform feature selection by assigning significant features larger weights. This process, however, consumes a large amount of time and requires a large amount of data. Removing less important factors can accelerate the training process of neural networks and reduce the amount of data needed. This paper therefore uses feature selection for all four types of models.

In statistics, a correlation coefficient is used to characterize how strong a relationship is between two variables. This coefficient is a real number between -1 and $+1$. The absolute value of correlation coefficient shows the strength of a relationship.

For example, assume there is a cyclist riding on a plane with constant velocity and direction. Assume there are two factors, one is the velocity of the cyclist, denoted as v, and the other is the color of the bike, denoted as c. Based on the property of uniform linear motion, the distance d can be described as the product of cyclist velocity v and time t.

To build a distance forecasting model, v rather than c should be considered. The correlation coefficient between v and d is $\rho_{v,d} = 1$ while that between c and d is $\rho_{c,d} = 0$, which shows there is a linear relationship between v and d, whereas c and d are uncorrelated. By calculating and comparing correlation coefficients between features and the target value, significant features can thus be extracted and irrelevant features can be filtered.

Two common correlation coefficients are widely used: the Pearson correlation coefficient and the Spearman's rank correlation coefficient. Pearson's correlation coefficient assesses linear correlation (Tutorials, 2018). In contrast, Spearman's rank correlation coefficient focuses on the monotonic relationship between two random variables (McDonald, 2014).

According to Bishara et al. (Bishara and Hittner, 2017), calculating Spearman's correlation coefficient for non-normal data may be an optimal strategy when the data size is larger than 20. For performance forecasting, the heart rate and velocity of a cyclist shows complex non-linear relationships with factors like cadence and the grade of the road when the total number of data items is far more than 20. Spearman's rank correlation coefficient is more suitable for calculating the correlation coefficient, so we therefore select this approach to filter the personalized and terrain factors compared with Pearson's correlation coefficient.

For Spearman's rank correlation coefficient, raw data (such as heart rate, speed, and the grade of course) should be converted to rank variables. Given a heart rate sequence HR, where hr is a heart rate in this sequence, the rank variable r_{hr} of hr is obtained as follows (Myers et al., 2010):

- Sort HR in ascending order and denote the sorted heart rate sequence as HR_s
- The position where hr is in HR_s is the rank variable r_{hr} of hr

For example, given a heart rate sequence over 4 seconds is $\{100, 102, 95, 86\}$, the corresponding rank value for this heart rate sequence is $\{3, 4, 2, 1\}$.

Assume there are two data sequences (e.g., X and Y) each of which has n examples. Based on the method described above, two ranked sequences rg_X

and rg_Y can be obtained. The formula for Spearman's correlation coefficient (Myers et al., 2010) is expressed in Formula 1,

$$\rho_{rg_X, rg_Y} = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X}\sigma_{rg_Y}} \tag{1}$$

where $cov(rg_X, rg_Y)$ is the covariance of ranked sequences rg_X and rg_Y, while σ_{rg_X} and σ_{rg_Y} are the standard deviations of the ranked sequences.

3.3 Overview of Our Dataset

For heart rate forecasting, the dataset used in this paper contains the grade of course, speed, heart rate, altitude, cadence, and distance at each second. We are interested in understanding how course-specific features impacted forecasting performance. We therefore use mountain biking data from trails in the Nashville, Tennessee, USA region.

Mountain biking courses have significant terrain-based variation, ranging from rocks and roots to very steep up-hill sections, as well as to mud. These variations in terrain lead to large variation in the physiological demands on the rider. These variations affect the muscles used, such as core and shoulder muscle engagement when riding over rocky terrain.

Our dataset was collected on a Ripmo AF mountain bike instrumented with a Garmin 830 biking computer connected to an accelerometer-based speed sensor mounted to the front hub of the bike. The bike's crank arms included a Quarq Sram XX1 Eagle Dub power meter that used embedded strain gauges to measure the power applied to the pedals within +/- 2%. The Quarq power meter directly measured the rider's pedaling cadence from accelerometers embedded in the crank arms. The Garmin 830 included GPS positioning and improved location tracking using a fusion of wheel rotation, speed, and GPS fix data. Finally, a Wahoo Tickr elctrode-based chest strap was used to measure heart rate and communicate the data to the Garmin 830.

For our study, we selected 8.71 miles of riding on a 10-mile courses in the Nashville area. The rider was a 40-year old male weighing approximately 210 pounds.[1] In cycling, functional threshold power (FTP) is commonly used to measure cyclist fitness. FTP estimates the maximum power that can be sustained by a cyclist for one hour. The FTP for the data collection cyclist was 240 measured using an indoor cycling trainer with a ramp testing protocol.

[1]Our current work focuses on a single rider to maximize understanding of how course-specific features could be learned since individual riders have been more thoroughly studied in prior work.

The dataset and all the source code we used to evaluate the machine learning and neural network models discussed in this paper is available from `github.com/EricXQiu/SportDataProcessing`.

3.4 Feature Selection

The first step in our dataset processing involved selecting features for model training. Before selecting these features their significance must be determined. Spearman's correlation coefficients (ρ) for each sequence are listed in Table 1. This table shows that the

Table 1: Spearman Correlation Coefficient Between Factors and Heart Rate

Features	Spearman correlation coefficient
speed	0.1826
grade of course	0.2524
cadence	0.2389
distance	0.0466
altitude	0.2586

grade of course, cadence, speed, and altitude significantly influence heart rate more than the other factors. We therefore selected these four factors as the features for our heart rate forecasting model.

We measured heart rate data with an electrode sensor on a chest strap. During cycling, the sensor might situate far apart from the skin due to motion, which can cause error on heart rate data. To reduce the error and smooth the heart rate, the heart rate sequence was therefore processed with a window average of 60 seconds. The grade of course, cadence, speed, and altitude features are also processed with the same window average process to ensure consistency with the heart rate data.

We split the dataset between a training set and test set. By convention, we used an 80% vs 20% train-test split ratio to split data into the training set and test set. For the LSTM-based model, the training set was the first 80% of the data and the test set was the remaining 20% rather than a random selection to account for the order dependence in heart rate data. To compare the results with LSTM-based models, the same train-test split process was also performed for the random forest model and neural networks.

4 HEART RATE FORECASTING MODEL COMPARISON

This section describes how we applied four machine learning models (random forest, feed forward neural networks, simple recurrent neural networks, and long short term memory) to build heart rate forecasting models that predict an cyclist's heart rate on a given course. The results of applying these heart rate forecasting models are then analyzed. We also compare the mean squared error (MSE) of all four models to glean insights into which models perform best and whether they have severe overfitting. Finally, we compare our models with other models (i.e., Ni's model (Ni et al., 2019) and Luo's model (Luo and Wu, 2020)) that are *not* course-specific, but which we trained on our dataset described in Section 3.3.

4.1 Random Forest Heart Rate Models

We used the scikit-learn library (Pedregosa et al., 2011) to construct the heart rate forecasting models based on random forest. The criterion is MSE and the number of decision tree regressors can be modified. Table 2 shows four random forest models with different numbers of decision trees regressors.

Table 2: Number of Decision Tree Regressors in RF model

Model Number	#Decision Tree Regressors
1	5
2	10
3	15
4	20

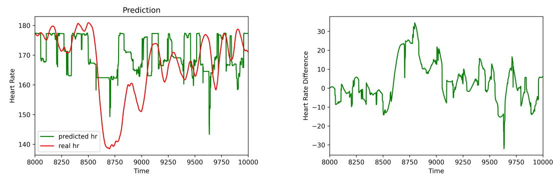
Each decision tree regressor uses all factors in regression. Their MSE on the training set and test set, as well as the maximum depth of the decision tree regressor, are shown in Table 3. For all four models,

Table 3: MSE and MaxDepth of RF Models on Data Set

RF Model	MSE on Training Set	MSE on Test Set	Max Depth
Model 1	0.455	125.15	31
Model 2	0.245	120.39	30
Model 3	0.157	130.40	34
Model 4	0.289	128.92	34

the MSE on the test set are hundreds of times larger than on the training set, which shows that these random forest models have severe overfitting. Table 3 shows the depth of decision tree regressors are approximately six times larger than the number of factors. This result indicates some factors are split multiple times, which increases the complexity of the decision tree, thereby yielding severe overfitting of the random forest model.

The heart rate forecasting result of model 1 is chosen and shown in Figure 2a. The predicted value follows the trend of the heart rate and shows the model learned some patterns in the heart rate sequence due to the course. However, the predicted heart rate shows a large error around time 8,900s. Moreover, the predicted heart rate remains stable for the time range from 8,600s to 8,700s, whereas the real heart rate shows a sharp decreasing trend. The error in percentage between predicted heart rate and real heart rate is shown in Figure 2b, which shows a similar trend as the forecasting result.



(a) Forecasting Result (b) Prediction Error

Figure 2: Heart Rate Forcasting of Random Forest Model 1

To investigate the structure of random forest model, the Gini importance[2] of each factor is calculated. A factor with large Gini importance means that more nodes are split by this feature in the decision tree and therefore this feature is considered significant.

The average Gini importance of factors in each model is listed in Table 4. This table shows that speed,

Table 4: Gini Importance of Random Forest Models

RF Model	Gini Importance of Factors (Time, Grade, Speed, Cadence, Altitude)
Model 1	0.294, 0.0376, 0.531, 0.028, 0.108
Model 2	0.280, 0.0344, 0.536, 0.029, 0.120
Model 3	0.281, 0.0342, 0.540, 0.030, 0.113
Model 4	0.285, 0.0349, 0.536, 0.028, 0.115

time, and altitude are three major factors used to split nodes in the first several layers of decision tree. Likewise, the table also shows the influence of cadence and grade are less important in decision tree construction.

We hypothesize that the source of the error may occur for the following reasons:

- The dataset only indirectly measures course features via speed and position, so it does not effectively learn course-specific influences on heart rate.
- There may be some internal rider conditions, such as the mental activities of the cyclist over time. For example, a cyclist may be anxious due to the

[2]The Gini importance is the average decrease of variance, which indicates the probability of whether there is a split on this feature (Menze et al., 2009).

difficulty of the terrain, thereby increasing their heart rate. All these features can influence heart rate, though they are hard to measure and express quantitatively.

- The random forest models may overfit on the training set and provide erroneous predictions.

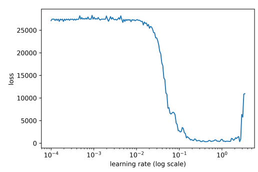
4.2 The FFNN Heart Rate Model

Due to the overfitting problems of the random forest models, we built feed forward neural network (FFNN) models with different hyper-parameters. The hyper parameters of the FFNN models are listed in Table 5.

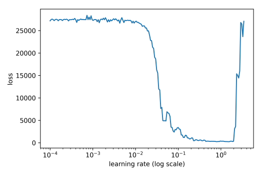
Table 5: Hyper-Parameters of FFNN Models

Model Number	#Layers	# Neurons in Each Layer
1	2	(5, 1)
2	2	(10, 1)
3	3	(5, 5, 1)
4	3	(10, 5, 1)

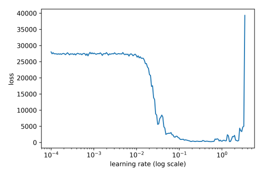
An important phase of training a neural network involves finding a proper learning rate. To search the learning rate systematically, we employed Smith's method (Smith, 2017) and utilized learning rate finder implemented by Pavel Surmenok et. al (Surmenok and Mackenzie, 2017). By running at each learning rate for 10 epochs, the loss vs learning rate curves are plotted, as shown in Figure 3. According to the



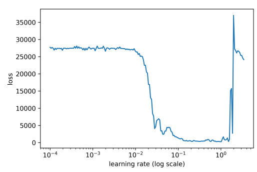
(a) Model 1



(b) Model 2



(c) Model 3

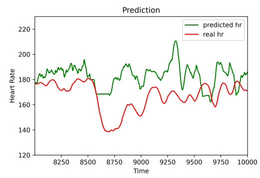


(d) Model 4

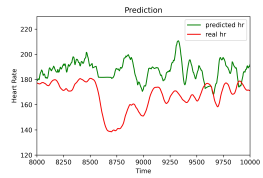
Figure 3: Learning Rate Search for FFNN Heart Rate Models

curves, the optimal start learning rate for these four FFNN models are 3×10^{-2}.

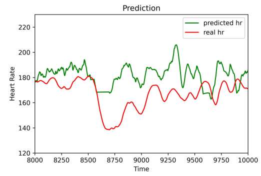
Figure 4 shows the predicted heart rate and the real heart rate of these four FFNN models applied to the same sample as the random forest. The sample shows



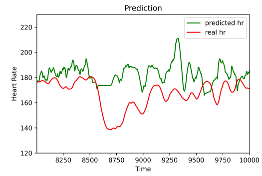
(a) Model 1



(b) Model 2



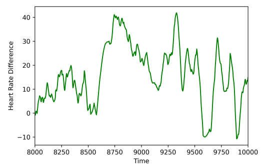
(c) Model 3



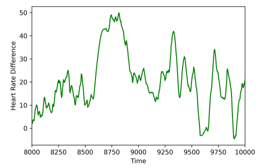
(d) Model 4

Figure 4: Heart Rate Forecasting of FFNN Models

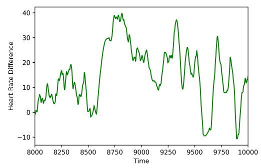
that FFNN models can follow the trend of the cyclist's heart rate on the specific course. The error and error percentage are plotted in Figure 5 and Figure 6, respectively.



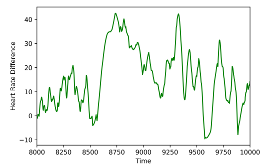
(a) Model 1



(b) Model 2



(c) Model 3



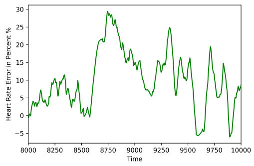
(d) Model 4

Figure 5: Heart Rate Forecasting Error of FFNN Models

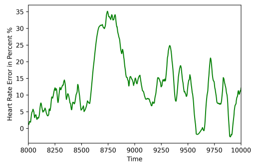
Models 3 and 4 show large errors from 8,500s to 9,000s. To investigate the error source, MSEs of the training set and test set are calculated for these four models, as shown in Table 6. Model 3 shows an obvious overfitting on the training set, while Models 1 and 2 show large bias on the training set. Among these four models, Model 4 shows relatively low bias and variance.

Table 6: MSE of FFNN models

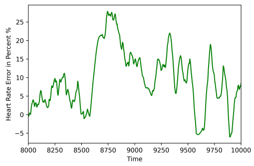
Model Number	MSE on Training Set	MSE on Test Set
1	242.89	405.11
2	238.81	568.08
3	262.47	364.91
4	233.25	402.34



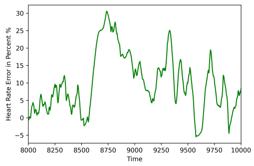
(a) Model 1



(b) Model 2



(c) Model 3



(d) Model 4

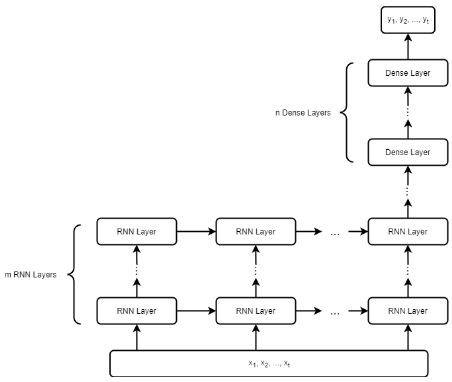
Figure 6: Heart Rate Forecasting Error in Percent of FFNN Models

Compared with random forest models, the variances of FFNN models on the training set and test set are much closer, indicating less overfitting severity. We therefore expect the FFNN models to generalize better than the random forest model because they do not incur such severe overfitting.

4.3 The Simple RNN Model

Unlike FFNNs, recurrent neural networks (RNNs) can utilize heart rate data in the past to forecast the heart rate at the current moment. Figure 7 shows the architecture of a simple RNN model for heart rate forecasting. The number of simple RNN layers, m, and the number of fully connected layers (also known as dense layers) n, are two important hyper-parameters in a simple RNN heart rate forecasting model. The heart rate sequence is divided by the maximum heart rate.

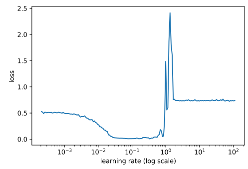
Figure 7: Structure of Simple RNN Heart Rate Forecasting Model



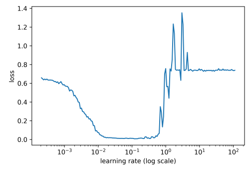
Simple RNN Models with different hyper parameters were built and their hyper-parameters are listed in Table 7. The optimal start learning rates were identified via Smith's method and an exponential decay learning rate function was utilized for learning rate searching. The learning rate vs loss curves are plotted in Figure 8. From the learning rate vs. loss curves shown in Figure 8, the optimal start learning rate was selected as 2×10^{-1}.

Table 7: Hyper-parameters of Simple RNN models

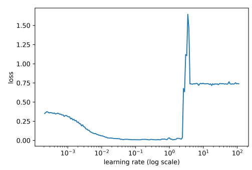
Model Number	#Dense Layers	#Simple RNN Layers
1	1	1
2	1	2
3	2	1
4	2	2



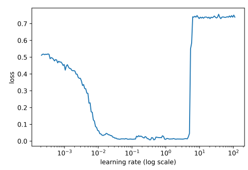
(a) Model 1



(b) Model 2



(c) Model 3



(d) Model 4

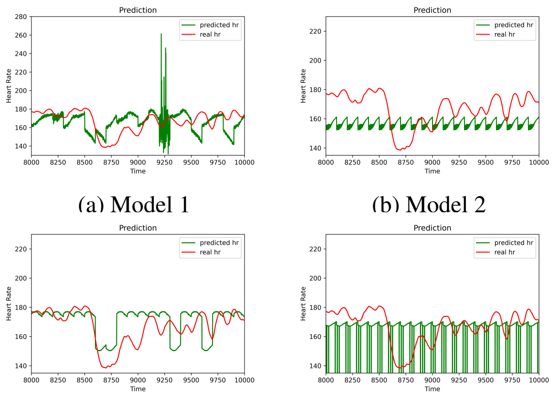
Figure 8: Learning Rate Search for Simple RNN Models

Table 8 shows the MSEs of four simple RNN models with different sets of hyper-parameters on the training set and test set. For Model 3, the MSE on the test set is much larger than that on the training set, which indicates overfitting.

Table 8: MSE of Simple RNN models

Model Number	MSE on Training Set	MSE on Test Set
1	275.11	248.83
2	628.48	237.68
3	210.31	655.18
4	1245.66	964.61

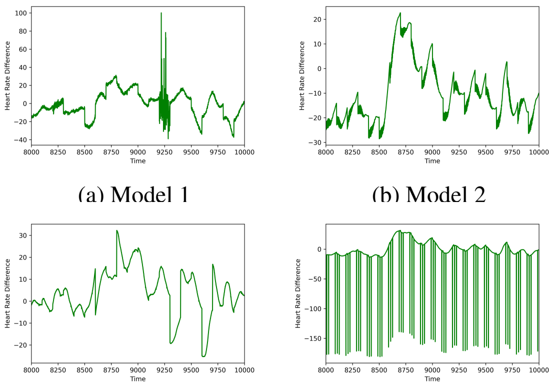
The heart rate of the athlete is predicted in Figure 9 and the error and error in percentage are shown in Figures 10 and 11. Compared with FFNN models, simple RNN Models 1 and 3 exhibit fewer errors and the forecasting results are closer to the real heart rate sequence.



(a) Model 1 (b) Model 2

(c) Model 3 (d) Model 4

Figure 9: Heart Rate Forecasting of Simple RNN Models



(a) Model 1 (b) Model 2

(c) Model 3 (d) Model 4

Figure 10: Heart Rate Forecasting Error of Simple RNN Models

4.4 The LSTM Model

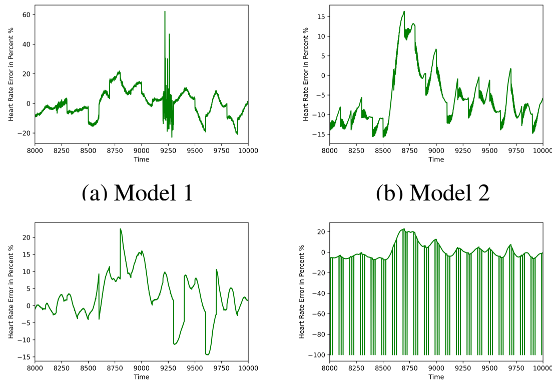
Simple RNN models can suffer from vanishing gradient problems, which limits their application when the sequence of input data items is very long. To address this issue, therefore, we also implemented LSTM models. The hyper-parameters of these models are shown in Table 9. The error of the results for

Table 9: Hyper-parameters of LSTM models

Model Number	#Dense Layers	#LSTM Layers
1	1	1
2	1	2
3	2	1
4	2	2

two-layer LSTM models is large, which likely occurs since models with two LSTM layers are so deep that our current dataset is insufficient to train them.

The learning rate vs. loss function curves that utilize the optimal learning rate finder are shown in Figure 12. The optimal start learning rate for models with only one LSTM layer (i.e., Models 1 and 3) is 1×10^{-1} and the optimal start learning rate for models with two LSTM layers (Models 2 and 4) is 1×10^{1}.

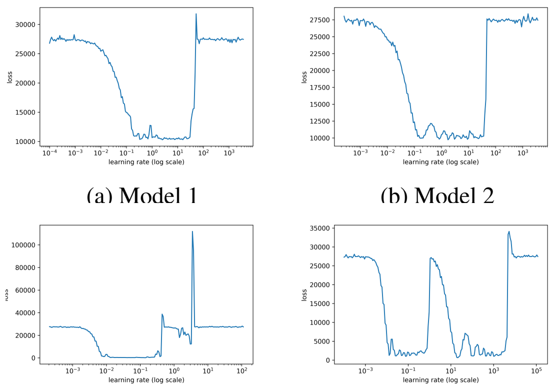


(a) Model 1 (b) Model 2

(c) Model 3 (d) Model 4

Figure 11: Heart Rate Forecasting Error in Percent of Simple RNN Models

The results for the four LSTM models with different hyper-parameters are shown in Figure 13. The heart rate error and error in percentage are shown in Figures 14 and 15, respectively.



(a) Model 1 (b) Model 2

(c) Model 3 (d) Model 4

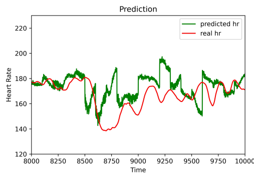
Figure 12: Learning Rate Search for LSTM Models

The MSE on the training set and test set is shown in Table 10. As shown in this table, the models with

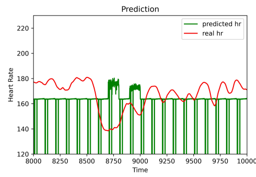
Table 10: MSE of LSTM models

Model Number	MSE on Training Set	MSE on Test Set
1	141.17	200.50
2	1240.60	1015.87
3	62.47	196.61
4	1480.36	1470.36

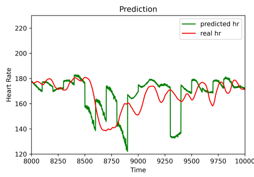
two LSTM layers exhibit large MSE over both the



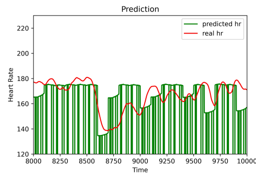
(a) Model 1



(b) Model 2

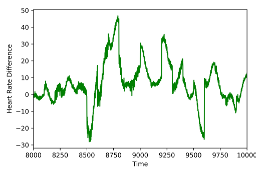


(c) Model 3

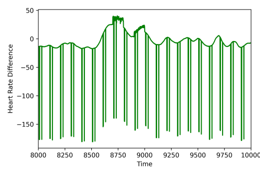


(d) Model 4

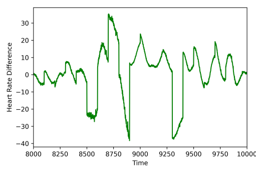
Figure 13: Heart Rate Forecasting of LSTM Models



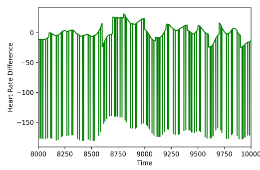
(a) Model 1



(b) Model 2



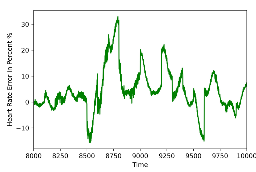
(c) Model 3



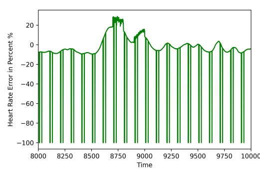
(d) Model 4

Figure 14: Heart Rate Forecasting Error of LSTM Models

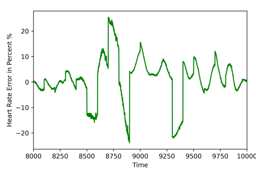
training set and test set. This larger error may occur since our dataset is insufficient to train a neural network with two LSTM layers. This result may also indicate that the simple one-dimensional signals of speed, position, and cadence are insufficient to learn



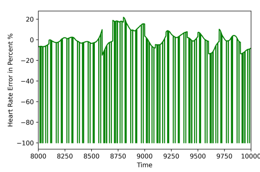
(a) Model 1



(b) Model 2



(c) Model 3



(d) Model 4

Figure 15: Heart Rate Forecasting Error in Percent of LSTM Models

course-specific heart rate variation.

4.5 Comparison with Other Heart Rate Models

Based on the results presented above, it appears that two LSTM models with only one LSTM layer perform better than the other three types of model. We therefore also compared these LSTM models with Luo's LSTM model (Luo and Wu, 2020) and Ni's LSTM-based model (Ni et al., 2019), which are course-independent and rely only on personalized factors and contextual factors. The input features to these latter two models were cadence, speed, altitude, and time.

Since Luo's and Ni's models are not course-specific, the grade of a biking course is excluded in the input factors. In particular, only the structure of their models are utilized and some layers (such as the encoding layers in Ni's model) are removed since heart rate forecasting is the main focus. The structure of these models are shown in Table 11.

Table 11: Structure of Heart Rate Forecasting Models

Models	Layer Structure of Models
Ni's model	LSTM + Dense + Dropout
Luo's model	LSTM + Dropout + LSTM + Dropout
LSTM model 1	LSTM + Dense
LSTM model 3	LSTM + Dense + Dense

These two models were first trained on the same training set as our heart rate forecasting models with course-specific factors excluded. They were then tested on the same test set. The MSE of all four of these models is shown in Table 12.

Table 12: MSE of LSTM models

Models	MSE on Training Set	MSE on Test Set
Ni's model	1822.23	407
Luo's model	239.89	157.40
LSTM model 1	141.17	200.50
LSTM model 3	62.47	196.61

The forecasted heart rate is shown in Figure 16. While Ni et al.'s model showed less error on both the training set and test set, the heart rate it forecasted is close to constant, except for several drop outs. Luo et al.'s model captures the heart rate pattern at around 170 beats per minute, but cannot follow the heart rate pattern overall.

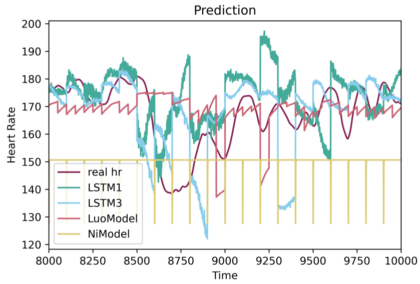


Figure 16: Comparison of Different Heart Rate Forecasting Models

5 SUMMARY OF OUR EXPERIMENT RESULTS

This section summarizes the results from our experiments described in Section 4.

5.1 Summary of Random Forest Model Results

The random forest models presented in Section 4.1 exhibit severe overfitting, likely because these models are based on decisions from an ensemble of tree regressors, which are data sensitive. In particular, tiny fluctuations in the input sequence can result in dramatically different outputs. Data fluctuation is common and unavoidable in real-world heart rate forecasting for cycling due to the mechanical vibration and the cyclist's movement. The Gini importance analysis shows that random forest models value speed, time, and altitude above grade and cadence, which is responsible for most of the errors. Due to severe overfitting, therefore, random forest models are poorly suited for personalized and course-specific heart rate forecasting in our dataset.

5.2 Summary of FFNN Model Results

Compared with the random forest models described above, feed forward neural networks (FFNNs) did not incur overfitting issues. Their major error source stems from the time dependency of heart rate, i.e., the heart rate at t-th second depends on what happened in the past (e.g., the cyclist rode up a steep hill). Errors also relate to the heart rate and other factors at $0, 1, \ldots, t-1$ seconds. Due to the nature of FFNNs, however, these models cannot deal with time dependencies efficiently.

5.3 Summary of the RNN Model Results

We tested two types of RNN models: a simple RNN model and an LSTM model. Our results showed that the simple RNN model captured the main trends of the heart rate sequence. The error source in the RNN models stemmed from either gradient vanishing or a gradient explosion, as described below:

- Gradient vanishing occurs in a long sequence, when the output value is large. In this case, the gradient of the RNN's sigmoid activation function is close to zero, leading to slow or no update to the weight matrices and bias vectors (Hochreiter, 1998).
- Gradient explosion occurs when an RNN is learning a long sequence and the gradient rises sharply, resulting in an unstable neural network.

These two problems may be the error source for simple RNN models, depending on sequence size.

In all four types of models, LSTM models show decent performance without severe overfitting. The LSTM has three gates that reduce the likelihood of vanishing or exploding gradients. For both simple RNN and LSTM models, however, models with two RNN/LSTM layers perform poorly since these two layer models are too deep to train efficiently with our dataset.

5.4 Mapping Our Results Onto Research Questions 1, 2, and 3

For research question 1 in Section 3.1 both course-independent models do not capture the heart rate trend of the cyclist well from 8,000s to 10,000s, as discussed in Section 4.5. In particular, the LSTM model proposed by Ni et al. (Ni et al., 2019) mainly outputs a rectangular heart rate sequence due to the lack of course-specific factors, such as grade. Ni et al. (Ni et al., 2019) state that their model focuses primarily on short-term predictions (typically in a window of 10 seconds), so a 2,000-second sequence may be too long for their model. Luo's model forecasted a somewhat continuous heart rate around 170 beats per minute with small dips around 9000s and 9250s. In general, our results show that these two models are not as accurate as our LSTM models since they do not consider course-specific factors.

For research question 2 in Section 3.1, random forest models exhibit lower MSE on the dataset. However, their substantial difference in MSE on the training set and test set indicates that they all incur severe overfitting. The maximum depth of each decision tree regressor in random forest models are all over six times the number of factors, which means that the random forest models split some factors over six times, which dramatically increases the complexity of the forecasting models and can result in severe overfitting.

Moreover, as the data size grows larger, the depth of the decision tree regressors also grows because the output decision tree regressor will cover all the heart rate cases, including any erroneous data in the dataset (e.g., due to sensor noise during collection). As the data set grows larger, the number of erroneous readings will unavoidably increase, thereby increasing the probability of erroneous prediction due to the accumulated influence of error. As a result, conventional machine learning models have difficulty on course-specific heart rate forecasting.

For research question 3 in Section 3.1 speed has the highest Gini importance, which aligns with prior work on heart rates for cyclists (Le et al., 2009) and is utilized as the splitting factor for the first layer. The three major factors are speed, time, and altitude. In contrast, the influence of cadence and grade are considered less important in decision tree construction.

In summary, our LSTM models 1 and 3 do not suffer from overfitting and offer reasonable heart rate forecasting. Compared with course-independent models, the forecasted heart rate of our course-specific models are closer to the real-world heart rate sequence.

6 CONCLUDING REMARKS

This paper presents an empirical analysis of personalized and course-specific models to forecast heart rates for cyclists. In particular, we explored the performance and feasibility of learning a cyclist's course-specific heart rate model from a single ride on a given course. We implemented models using long short term memory (LSTM), recursive neural network (RNN), random forest, and previously published architectures and compared their performance learning a cyclist. We also compared our model with Ni et al's (Ni et al., 2019) and Luo et al.'s (Luo and Wu, 2020) models.

The following are key lessons learned from our research on single-ride personalized and course-specific heart rate forecasting in cycling:

- **Overfitting is a challenge for traditional machine learning models.** Our MSE analysis in Section 4.1 showed the random forest models had severe overfitting, due to the effect of erroneous data (e.g., noise in commodity sensors) in the training set. Likewise, as the data set grew, erroneous readings introduced poor performance in the random forest forecasting models.
- **Course-specific factors are crucial in heart rate forecasting.** In Section 3.4, the Spearman's correlation coefficient showed that the grade of the course (which is a course-specific factor) was highly correlated with heart rate. Moreover, Section 4.5 showed that two course-independent models captured heart rate patterns poorly compared with our course-specific LSTM models.
- **Course-specific models offer accurate heart rate forecasting.** The results from our comparisons indicated that our LSTM-based models exhibit slightly lower mean square error (MSE) and mean absolute percentage error (MAPE) compared with Luo et al.'s model. Likewise, Jianmu Ni et al's model does not offer reasonable heart rate forecasting on the given dataset compared with our LSTM-based models.
- **Personalized and course-specific LSTM models can be learned for a cyclist to forecast heart rate from a single ride of a course.** More work is needed to validate this observation, but our initial results are promising. The heart rate forecasting results in Section 4 showed that the accuracy of the LSTM models outperformed the other two neural network models and did not have as severe overfitting as the random forest models. In Section 4.5, the LSTM course-specific models also performed better than the two LSTM-based course-independent models.

Our future work focuses on scaling up our validation on a larger body of cyclists to determine whether these results hold true across a range of riders. We are also exploring how imagery of the course can aid in understanding complicated course features, such as terrain roughness. For personalized factors, we are evaluating the extent to which learned course-specific models transfer to other riders of the same gender and age, as well as bike types. We are also considering dynamic personalized factors, such as breathing rate.

For course-specific factors, image data and videos are being collected and analyzed via neural networks. We are analyzing roughness and course conditions at different parts of the course from these images and applying them in our heart rate forecasting model.

REFERENCES

Bian, M., Peng, B., Wang, W., and Dong, J. (2019). An accurate LSTM based video heart rate estimation method. In *Lecture Notes in Computer Science (in-*

cluding subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 11859 LNCS, pages 409–417. Springer International Publishing.

Bishara, A. J. and Hittner, J. B. (2017). Confidence intervals for correlations when data are not normal. *Behavior Research Methods*, 49(1):294–309.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pages 1724–1734.

Chowdhury, S. S., Hasan, M. S., and Sharmin, R. (2019). Robust Heart Rate Estimation from PPG Signals with Intense Motion Artifacts using Cascade of Adaptive Filter and Recurrent Neural Network. In *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, volume 2019-October, pages 1952–1957.

Collinson, P. O., Boa, F. G., and Gaze, D. C. (2001). Measurement of Cardiac Troponins. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, 38(5):423–449.

Dawson, E. A., Shave, R., George, K., Whyte, G., Ball, D., Gaze, D., and Collinson, P. (2005). Cardiac drift during prolonged exercise with echocardiographic evidence of reduced diastolic function of the heart. *European Journal of Applied Physiology*, 94(3):305–309.

Gleeson, M. (2002). Biochemical and immunological markers of overtraining. *Journal of Sports Science and Medicine*, 1(2):31–41.

Hochreiter, S. (1998). Recurrent neural net learning and vanishing gradient. *International Journal Of Uncertainity, Fuzziness and Knowledge-Based Systems*, 6(2):107–116.

Jeukendrup, A. and Diemen, A. V. (1998). Journal of Sports Sciences Heart rate monitoring during training and competition in cyclists Heart rate monitoring during training and competition in cyclists. *Journal of Sports Sciences*, 16(1):91–99.

Le, A., Jaitner, T., Tobias, F., and Litz, L. (2009). A Dynamic Heart Rate Prediction Model for Training Optimization in Cycling (P83). *The Engineering of Sport 7*, pages 425–433.

Leijnen, S. and van Veen, F. (2020). The Neural Network Zoo. *Proceedings*, 47(1):9.

Lucía, A., Hoyos, J., and Chicharro, J. L. (2001). Preferred pedalling cadence in professional cycling. *Medicine and Science in Sports and Exercise*, 33(8):1361–1366.

Luo, M. and Wu, K. (2020). Heart rate prediction model based on neural network. *IOP Conference Series: Materials Science and Engineering*, 715(1):12060.

McDonald, J. H. (2014). *Handbook of Biological Statistics - Paired t–test*, volume 2. sparky house publishing Baltimore, MD.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., and Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*, 10(1):213.

Ming, Y. and Jun, J. (2008). Heart rate prediction based on physical activity using feedforward neural network. *Proceedings - 2008 International Conference on Convergence and Hybrid Information Technology, ICHIT 2008*, pages 344–350.

Mohammadzadeh, F., Nam, C. S., and Lobaton, E. (2018). Prediction of Physiological Response over Varying Forecast Lengths with a Wearable Health Monitoring Platform. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July:437–440.

Myers, J. L., Well, A., and Lorch, R. F. (2010). *Research Design and Statistical Analysis - Jerome L. Myers, Arnold Well, Robert Frederick Lorch - Google Books*. Routledge.

Ni, J., Muhlstein, L., and McAuley, J. (2019). Modeling heart rate and activity data for personalized fitness recommendation. *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 2:1343–1353.

Nieman, D. C. (1994). Exercise, infection, and immunity. *International journal of sports medicine*, 15(3):S131.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830.

Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*, pages 464–472.

Surmenok, P. and Mackenzie, J. (2017). keras_lr_finder. *GitHub repository*.

Tutorials, S. (2018). Pearson Correlation. *A Practical Approach to Using Statistics in Health Research*, 4:165–172.

Xiao, F., Chen, Y. M., Yuchi, M., Ding, M. Y., and Jo, J. (2010). Heart rate prediction model based on physical activities using evolutionary neural network. *Proceedings - 4th International Conference on Genetic and Evolutionary Computing, ICGEC 2010*, pages 198–201.