Construction of a Machine Learning Model to Estimate Physiological Variables of Speed Skating Athletes Under Hypoxic Training Conditions

Junhao Han,[1] Mingyang Liu,[1] Jizu Shi,[2] and Yuguang Li[1]

[1]College of Instrumentation and Electrical Engineering, Jilin University, Jilin Province, China; and [2]Jilin Institute of Physical Education, Jilin Province, China

Abstract

Han, J, Liu, M, Shi, J, and Li, Y. Construction of a machine learning model to estimate physiological variables of speed skating athletes under hypoxic training conditions. *J Strength Cond Res* XX(X): 000–000, 2021—Monitoring changes in athletes' physiological variables is essential to create a safe and effective hypoxic training plan for speed skating athletes. This research aims to develop a machine learning estimation model to estimate physiological variables of athletes under hypoxic training conditions based on their physiological measurements collected at sea level. The research team recruited 64 professional speed skating athletes to participate in a 10-week training program, including 3 weeks of sea-level training, followed by 4 weeks of hypoxic training and then a 3-week sea-level recovery period. We measured several physiological variables that could reflect the athletes' oxygen transport capacity in the first 7 weeks, including red blood cell (RBC) count and hemoglobin (Hb) concentration. The physiological variables were measured once a week and then modeled as a mathematical model to estimate measurements' changes using the maximum likelihood method. The mathematical model was then used to construct a machine learning model. Furthermore, the original data (measured once per week) were used to construct a polynomial model using curve fitting. We calculated and compared the mean absolute error between estimated values of the 2 models and measured values. Our results show that the machine learning model estimated RBC count and Hb concentration accurately. The errors of the estimated values were within 5% of the measured values. Compared with the curve fitting polynomial model, the accuracy of the machine learning model in estimating hypoxic training's physiological variables is higher. This study successfully constructed a machine learning model that used physiological variables measured at the sea level to estimate the physiological variables during hypoxic training.

Key Words: mathematical model, RBC, Hb, LSTM, data analysis

Introduction

Hypoxic training has been widely used in athletes' training because it can improve training efficiency (15). Many studies have investigated hypoxic training's effect and feasibility in different sports under different conditions (4,18,27). Many studies have shown that appropriate hypoxic training can improve physiological variables, such as red blood cells (RBCs) and hemoglobin (Hb), which are associated with athletes' performance in some sports. Hypoxic training is necessary for elite athletes if executed and monitored properly (15,28). However, hypoxic training beyond the athletes' endurance limit will not benefit the athletes (23). Instead, excessive hypoxic training may even affect the athletes' health and increase the risk of hypoxemia (8,19,26) or heart disease in athletes (2). Thus, it is crucial to ensure hypoxic training safety and maximize the training effect by creating an appropriate training environment and determining a proper training intensity. Some variables related to oxygen transport capacity can reflect the impact of hypoxic training on aerobic endurance performance. Changes in these variables can be used as references to construct a hypoxic training plan.

From a technical point of view, machine learning–based acquisition models have attracted extensive attention in sports (5). From a machine learning perspective, athletes' physiological variables can be used to evaluate their aerobic endurance (27). Hypoxic training can improve sports performance and achieve the best competitive state for athletes before the competition (4,17). Hypoxic training plans are generally formulated several months in advance. If the athletes' physiological variables after completion of hypoxic training can be estimated in advance, it will help develop hypoxic training plans. Machine learning methods can extract model to estimate the athletes' physiological variables after participating in hypoxic training. The model can estimate the physiological variables after completion of hypoxic training based on the physiological variables of athletes measured during sea-level training conditions. According to the estimated values, the coach can provide a hypoxic training schedule and exercise guidelines to athletes or seeded players in the same sport and altitude before the training (10).

From a safety point of view, a lower oxygen concentration level may not be associated with a better training effect. For example, 1 study compared the hypoxic training effect between athletes trained in an 1,800-m environment and 2,400-m environment, respectively (20). They found that the 2 groups of athletes showed a comparable increase in their blood Hb concentration level despite the difference in training altitude. In traditional hypoxic training, the coach constructed the training plan based on their previous experience (20). In a traditional hypoxic training plan, a coach will ask the athletes to adapt to a hypoxic training environment within about a week and gradually change the training

Address correspondence to Mingyang Liu, 740118972@qq.com.

Journal of Strength and Conditioning Research 00(00)/1–9

schedule and duration. This approach is to adjust the hypoxic load on the athletes progressively and to avoid excessive hypoxic training. The athletes can drop out from the training at any time when they experience any discomfort. This approach is secure but less efficient, and it relies heavily on the experience of the hypoxic training coach. If each athlete evaluates the changes in athletes' physiological variables after completion of hypoxic training through machine learning methods before starting hypoxic training, efficiency and safety will be improved (21).

"Living high-training low (HiLo)" is a commonly used method in hypoxic training, in which athletes live in high altitudes with low oxygen concentration levels, and train in low altitudes with normal oxygen concentration levels (6). Compared with traditional hypoxic training with intermittent hypoxic exposure, this method has been shown to improve aerobic endurance (3) and reduce side-effects in hypoxic training for athletes (11,16). Thus, this study adopted this hypoxic training procedure.

This article proposes a machine learning method to estimate the changes of physiological variables of athletes during hypoxic training. In this study, first of all, the physiological variables in a group of speed skating athletes were measured at sea level and 2,300-m simulated altitude environment. Second, a mathematical model was established to estimate changes in the physiological variables. This mathematical model was combined with the machine learning method to build a prediction model to estimate physiological variables during hypoxic training using data collected at sea-level training. The experimental results show that by extrapolating from the changes in athletes' physiological variables before and during hypoxic training, this method successfully estimated the average values of physiological variables among other athletes participating in this experiment during hypoxic training based only on measurements taken before hypoxic training.

Methods

Experimental Approach to the Problem

Figure 1 shows the main technology road map of this study, which includes 3 parts. First of all, we selected high-performance level athletes to participate in hypoxic training and measured their physiological variables. Second, mathematical modeling was built to reflect the physiological variables changes of athletes according to the collected data. Based on the mathematical model, we get information on the physiological variable changes. Finally, the obtained mathematical model was processed and then used in machine learning. Processing includes acquiring the sequence (discretization) from the mathematical model and dividing the sequence into a sea-level training section and a hypoxic training section. The purpose of processing was to improve the training efficiency and performance of machine learning models. The processed data were used for deep learning through the long short-term memory (LSTM) network until an estimation model with a root-mean-square error (RMSE) less than 0.001 was obtained.

The experimental protocol lasted for 10 weeks, including the first 3 weeks before hypoxic training (pre-HiLo), the 4-week hypoxic training (HiLo), and the last 3 weeks of recovery (post-HiLo). Data from the final 3 weeks were used to observe any changes in athletes' physiological variables and any adverse effects on athletes after the hypoxic training. The Hypoxic Tent System (SenLuo, China) manufactured by Tianjin Senluo company was used to create a hypoxic training environment. The oxygen concentration in the cabin was 15.8–16.1% (equivalent to the altitude between 2,300 and 2,500 m), the temperature was 18–23° C, and the air humidity was 30–52%. Athletes trained daily at sea level during the day and rest in the cabin with low oxygen concentration level. Athletes left the cabin at 6:00 AM and returned at 10:00 PM and slept for at least 8 hours every day for 4 consecutive weeks. Figure 2 shows the cabin that was used to create a hypoxic training environment for athletes.

Subjects

A total of 64 subjects participated in the study, including 32 women (age = 20.68 ± 3.41 years, body height = 171.00 ± 4.60 cm, body mass = 60.1 ± 2.49 kg, and training years = 7.60 ± 1.41) and 32 men (age = 19.85 ± 4.35 years, body height = 178.86 ± 5.51 cm, body mass = 65.64 ± 5.72 kg, and training years = 8.71 ± 2.05). All the subject information (including age, height, weight, etc.) are measured as Â ± SE. All athletes were healthy national-level speed skating athletes with no respiratory, cardiovascular, or endocrine diseases. All subjects had no history of smoking or drinking. They participated in physical training every year regularly and had not undergone any hypoxic training sessions in the past year. All athletes signed the written informed consent forms before participating. All athletes were in breaks and had no competition tasks. All experimental protocols were conducted following relevant guidelines and regulations. This study was performed by College of Instrumentation and Electrical Engineering, Jilin University, Jilin Province, China, and assisted by Jilin Institute of Physical Education.

Procedures

Blood oxygen transport capacity is the maximum amount of oxygen the blood can transport. RBC count and Hb concentration are essential parameters to estimate oxygen transport capacity. Increasing oxygen transport capacity could also improve aerobic endurance performance (29). Hypoxic training affects the athletes' aerobic endurance performance by affecting oxygen transport capacity (20,30). In hypoxic training, variables associated with oxygen transport capacity are often used to indicate the effect of hypoxic training. Monitoring the changes of these parameters is also crucial for the safety of hypoxic training. Some studies have shown that a decrease in blood oxygen saturation leads to an increase in Hb level, counteracting the adverse effects of a low blood oxygen level (24). The primary function of Hb is to transport oxygen molecules from lung to tissues. Oxygen molecules delivered to muscles will be used to produce adenosine triphosphate, which is the energy source for muscle activation during exercise. When the oxygen concentration in the environment decreases, the body's relative oxygen uptake decreases, resulting in a reduced oxygen binding capacity. In response to this condition, the body increases the RBC count and Hb concentration to maintain a consistent blood oxygen transport capacity (15).

Therefore, RBC count and Hb concentration were selected as the primary physiological variables to reflect the hypoxic training effect. We used pocH-100I Automatic Tri-classification Blood Analyzer (Sysmex, Germany) to measure RBC count and Hb concentration in the experiment. The blood tests were conducted in the morning of the last day of every week during training. The athletes were sampled in a fasting state, and they were asked to lie in bed for 5 minutes before the blood test. We collected each blood sample at the same time every week to account for potential day-to-day variance.

Establishment of Mathematical Models

Physiological variables were measured by taking blood samples from athletes. However, the physiological variables measured

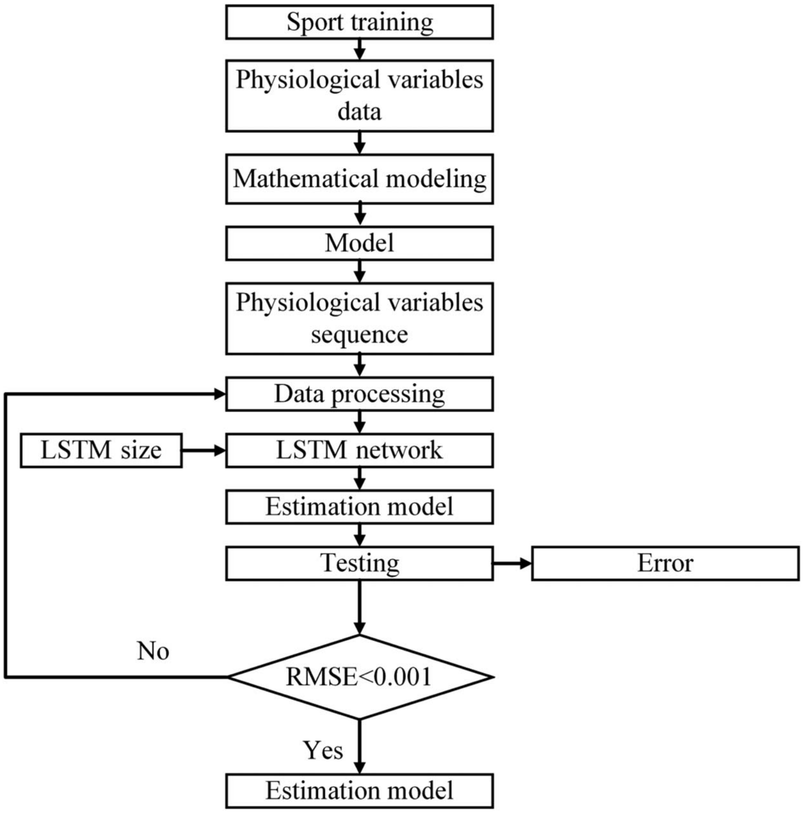

Figure 1. The technology roadmap for machine learning to estimate physiological variables.

once a week by blood samples can only provide limited information because of the measurement frequency. This study established a mathematical model to simulate athletes' physiological variables and obtained a curve of physiological variables changing with time-based on the data's physiological variables. The changes of the physiological variables across the whole experiment include 3 parts, including the sea-level training period, transition period from sea level to hypoxic training, and stabilized period when physiological variables adapt to the lower oxygen concentration level.

Human physiological variables conform to Gaussian distribution (7). The physiological variables measured in the experiment can be treated as sample points of random signals. The maximum likelihood method can estimate the random signal's overall situation based on samples with known probability density of data (8,13). The physiological variables of the human body changes over time can be treated as a continuous variable X. The sample from X is X_i. The probability density of X can be expressed as $f(x;\theta)$, where x is the observed value of the sample, and θ is the probability density function parameter.

During the training at sea level, the athletes were measured 3 times. The samples of measurements were marked as X_1, X_2, and X_3, and correspondingly, the samples' observed values were x_1, x_2, and x_3. The probability that a random point X_i falls in the point x_i neighborhood with a length ofΔx_iapproximately equals$f(x_i;\theta)\cdot\Delta x_i$. The likelihood of the sample approximately equals to $\prod_{i=1}^{3} f(x_i;\theta)\cdot\Delta x_i$. According to the maximum likelihood estimation method, $\prod_{i=1}^{3} f(x_i;\theta)\cdot\Delta x_i$should achieve the maximum value at an estimated value $\theta$of parameter θ. Therefore, the likelihood function $L(\theta)$ of physiological variables is

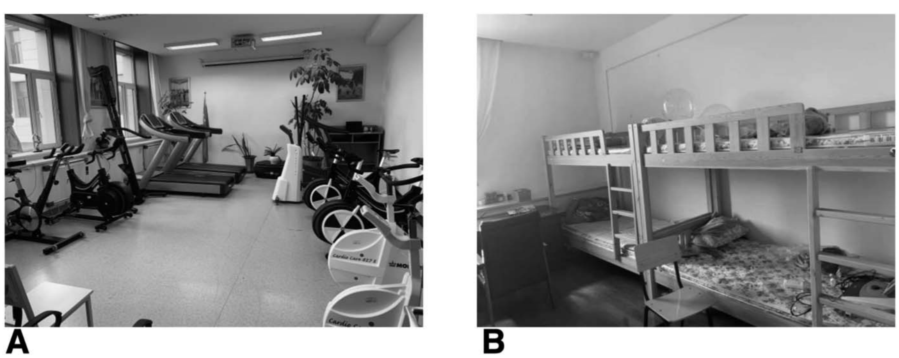

Figure 2. The system (A) used for athlete training and the hypoxic tent system (B) used for athlete rest.

$$L(\theta) = \prod_{i=1}^{3} f(x_i;\theta)\cdot\Delta x_i \tag{1}$$

For a Gaussian distribution of a single variable, its probability density function can be expressed. Therefore, we can get the likelihood function of physiological variables. The estimation function of the sample population can be determined by the formula (2). μ is the mean of the sample, and σ is the variance of the sample.

$$\begin{cases} \dfrac{\partial \mathrm{In}L}{\partial \mu} = \dfrac{1}{\sigma^2}\left(\sum\limits_{i=1}^{3} x_i - n\mu\right) = 0 \\ \dfrac{\partial \mathrm{In}L}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{3}(x_i-\mu)^2 = 0 \end{cases} \tag{2}$$

According to the sample population's estimation function, we can determine the population trend of sea-level training physiological variables. Similarly, this method can also be applied to the mathematical model of physiological variables after entering the hypoxic training plateau.

A previous study assessed the changes in physiological variables during the transition period between training at the sea level and hypoxic training (25). This study reported that the Hb, HA, and other physiological variables related to oxygen transport capacity usually increase at a fast constant rate during the first 2 to 3 weeks of training. Then, the increase rate begins to decline until reaching a plateau, and lastly, keep steady (25). In our article, we used a polynomial function to represent the mathematical model of physiological variables' changes during the transition from sea level to a hypoxic environment (1,22). The athletes' variables before entering hypoxic training and the athletes' variables in the first 2 weeks during hypoxic training were selected as the reference to get the mathematical model. The high-order polynomial function obtained by curve fitting can describe the changes in physiological variables of hypoxic training in the transition period to some extent.

Therefore, the variables measured during the experimental process can be transformed into the above mathematical model. Figure 3 compares the directly measured RBC counts and the mathematical model's values established according to the above methods. Compared with directly measuring the change of the original data during the hypoxic training, this mathematical model has the following advantages: First of all, the mathematical model shows a continuous change in the physiological variables during training, which offers more information to train the machine learning model. Secondly, when directly measured data are used to describe the changes of physiological variables, there is a certain degree of uncertainty. If a specific measurement has a large deviation from the group average, it will significantly impact the general prediction results. The mathematical model can better reflect the overall change of physiological variables and reduce an outlying datapoint's impact in a particular section.

Establishment of the Machine Learning Model to Estimate Physiological Variables

Two main functions of machine learning are classification and prediction. As a machine learning method, a recurrent neural network (RNN) is primarily used to process time series and extract information about how they change (12,31). People can obtain the previously described mathematical model by modeling the changes in physiological variables over time, which can be converted to a time series. Therefore, people can use RNN to estimate physiological variables such as RBC count and Hb concentration.

According to our mathematical model of physiological variables, each athlete's physiological variables can be transformed into a sequence of changes over time. The sequence is denoted as$X_i = [x_1, x_2, \cdots, x_\mathrm{n}]$. In the experiment, a series of multiple physiological variable sequences can be obtained by the mathematical model. For a set of measurements, the sequence is$X_i = [x_1, x_2, \cdots, x_\mathrm{n}]$, where n was the length of the sequence. For a group of measurement results, the first k values in a sequence of physiological variable measurements belonging to the sea-level training part of the sequence were taken as input. The first k values in a sequence refer to the part of sea-level training in the sequence transformed from the mathematical model. The average value of m values in a sequence after the physiological variable measurements during hypoxic training stabilized was taken as output. The m values refer to the partial sequence reflecting physiological variables' changes during the plateau period of hypoxic training. Cross-validation is used in this study to assess a machine learning model's accuracy (14).

Machine learning requires multiple rounds of training to obtain a model that could estimate physiological variables accurately. Long short-term memory network is a particular RNN,

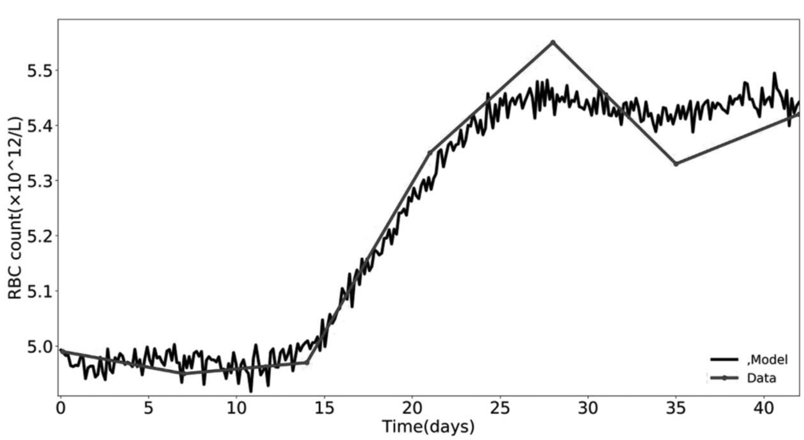

Figure 3. Comparison of the measured RBC count data and the curve of the mathematical model. RBC = red blood cell.

which performs better than traditional RNN in long sequence learning (9). The estimation model of physiological variables obtained through machine learning is based on the LSTM network. For the sequence, $X_i = [x_1, x_2, \cdots, x_n]$changes with time, where x_t is the value any time in the sequence, and the variable physiological information is updated according to LSTM. Therefore, the input gate function is:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_t), \tag{3}$$

$$C = \tanh(W_C \cdot [h_{t-1}, x_t] + b_t). \tag{4}$$

σ is the activation function, which is a rectified linear unit in this study. W_i and W_C are the input gate's weight matrices, and b_t and b_C are the bias vectors. The input gate will record the feature information of the current input variable x_t in the LSTM network as the network's physiological variables are updated.The function of forget gate is

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right). \tag{5}$$

$W_f$is the weight matrix of the forgotten gate, and b_f is the bias vector of the forgot gate. Changes in the physiological variables are continuous. A forget gate is used to eliminate some unnecessary information selectively to avoid excessive effects from the previous input on the current input physiological variables.-Status update function

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t. \tag{6}$$

The status update function is used to update information about the physiological variables currently stored in the LSTM network.

The output gate function is:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{7}$$

$$h_t = o_t \cdot \tanh(C_t). \tag{8}$$

In the LSTM iterative network, $C_{t-1}$and $h_{t\text{-}1}$ are the updated state and output of the network at the time *t-1*. A model *F(X)* will be obtained after all physiological variable sequences passing through the LSTM.

Long short-term memory training networks can adjust parameters according to different data types and data volumes to achieve better results. In the LSTM network, LSTM size is used to describe the size of the RNN network in the above-mentioned training process. A larger LSTM size means more complex the training network used in training. More complex networks can improve the accuracy of the estimates. In the study, we used different LSTM sizes to construct different machine learning models and compared the prediction accuracy among these models.

All the methods described in this section are implemented based on TensorFlow by Python. In TensorFlow, there are built-in frameworks for the implementation of machine learning and LSTM. For athletes of a particular sport, use this method to estimate the athletes' physiological variables in a specific hypoxic environment. First, collect about 100 sets of data that have completed hypoxic training in this environment. The data need to be collected at sea level for more than 2 weeks, collecting for more than 4 weeks in hypoxic training, and the measurement frequency is not less than once a week. Select physiological variables need to be estimated (such as red blood cells, Hb, blood oxygen saturation, etc.). Then, each set of data need to be converted into a corresponding mathematical model. Build an RNN network framework based on the TensorFlow framework, try different LSTM sizes, and select the best LSTM size based on the trial results. Use this LSTM size for machine learning, and train a machine learning model. The sea-level physiological variables of athletes who need to estimate hypoxic training's effect are transformed into mathematical models and input into the trained machine learning model. Finally, the physiological variables of athletes undergoing hypoxic training in this particular hypoxic environment can be estimated.

Table 1

Physiological variables of men in the study after hypoxic training.

	Mean value (sea level)	Mean value (hypoxia)	Increasing rate
RBC ($\times 10^{12}$/L)	5.060 (±0.70)	5.260 (±0.85)	3.86%
Hemoglobin (g/dl)	15.27 (±1.53)	16.13 (±1.67)	5.11%

RBC = red blood cell.

Results

Statistical Results of Hypoxic Training

After the hypoxic training, the athletes recovered well without any adverse symptoms. The physiological effects of hypoxic training continued for at least a week after the athletes leave the cabin. We recorded the mean and the *SD* of RBC count and Hb concentration values of the male (Table 1) and female (Table 2) athletes in hypoxic training. In Tables 1 and 2, the increasing rate displays the average increasing rate of RBC count and Hb concentration of each athlete after completion of hypoxic training. After hypoxic training, physiological variables usually increase by more than 5%, as reported in previous studies, which has been considered a considerable improvement (30). In this study, the average amount of increase in physiological variables was about 4% (male) and 9% (female), and 51 of the 64 athletes increased their RBC count and Hb concentration by more than 5% after completion of hypoxic training.

Figure 4 shows the changes in the average physiological variables of all athletes over time. We recorded the RBC count and Hb concentration in the experiment. The black broken lines in the figure represent the mean and *SD* of the measured physiological variables from male athletes. The blue lines represent the physiological variables' mean and *SD* of female athletes.

Machine Learning Predicts Results

Table 3 is the experimental results of physiological variables estimated by machine learning using actual measured data. This table shows the RMSE of the predictive machine learning model of RBC count and Hb concentration of the athletes after 1 to 110 rounds of training under different LSTM sizes.

Figure 5 shows the changes in RMSE values versus the number of training rounds in male and female athletes under different

Table 2

Physiological variables of women in the study after hypoxic training.

	Mean value (sea level)	Mean value (hypoxia)	Increasing rate
RBC ($\times 10^{12}$/L)	4.449 (±0.64)	4.841 (±0.94)	9.34%
Hemoglobin (g/dl)	13.09 (±1.61)	14.391 (±2.31)	9.73%

RBC = red blood cell.

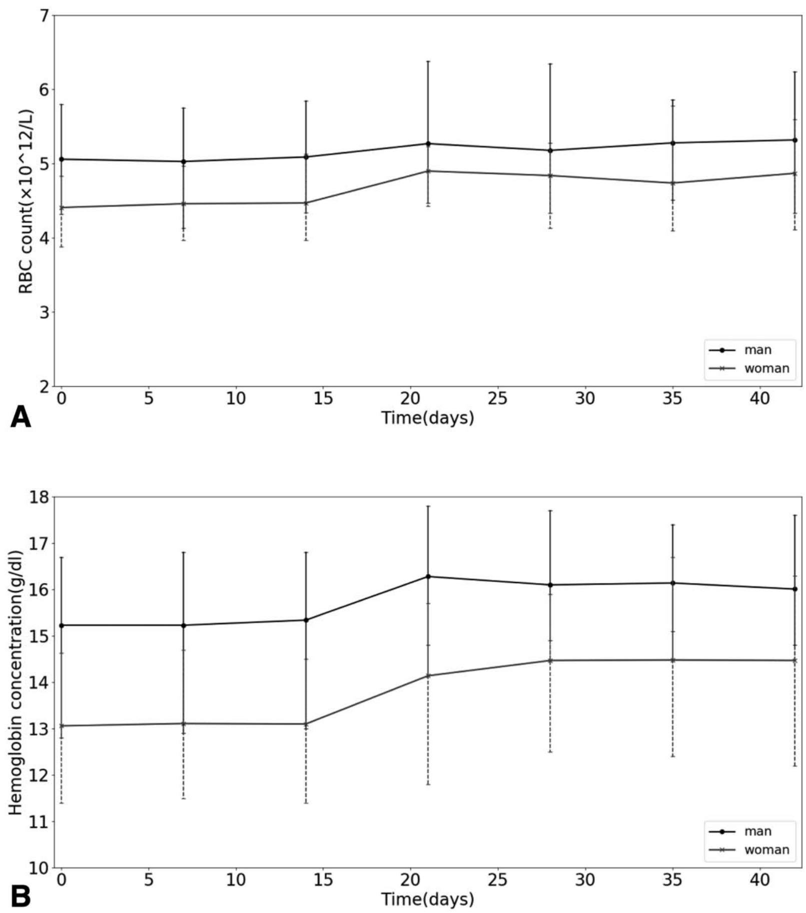

Figure 4. Changes of physiological variables of athletes across the 3-week training at sea-level and the 4-week hypoxic training. A) RBC count over time and (B) Hb concentration over time. Hb = hemoglobin; RBC = red blood cell.

Table 3

Performance of machine learning in the estimation of physiological variables.

Number of training rounds	RMSE (120)	RMSE (240)	RMSE (480)	RMSE (720)	RMSE (960)
Female					
1					
RBC	0.059939	0.087018	0.100513	0.026249	0.039148
Hb	0.055760	0.043775	0.066601	0.027490	0.056321
60					
RBC	0.022866	0.022796	0.022359	0.013183	0.014187
Hb	0.022467	0.019920	0.018595	0.014500	0.016952
110					
RBC	0.020301	0.018423	0.016308	**0.010656**	**0.010641**
Hb	0.020317	0.016543	0.015224	**0.009915**	**0.010619**
Male					
1					
RBC	0.054913	0.591007	0.665052	0.033858	0.023027
Hb	0.093997	0.023192	0.019710	0.025732	0.021909
60					
RBC	0.020328	0.118302	0.097890	0.013350	0.012316
Hb	0.018450	0.010563	0.009654	0.009983	0.009417
110					
RBC	0.017329	0.094820	0.086107	**0.010556**	**0.009456**
Hb	0.015566	0.008907	0.008554	**0.008339**	**0.007436**

Hb = hemoglobin; RBC = red blood cell; RMSE = root-mean-square error.

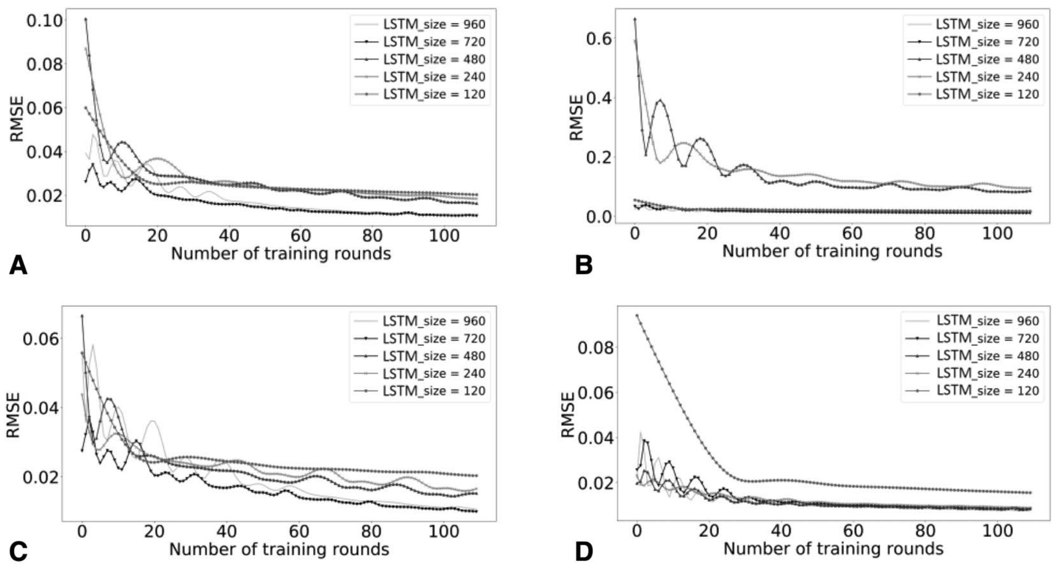

Figure 5. RMSE values of the predictive machine learning model versus the number of training rounds under different LSTM sizes. Estimation of RBC count of female athletes (A) and male athletes (B). Estimation of hemoglobin concentration of female athletes (C) and male athletes (D). LSTM = long short-term memory; RMSE = root-mean-square error; RBC = red blood cell.

LSTM sizes. All machine learning models with different LSTM sizes converged in the end.

Table 4 lists the accuracies of the machine learning model and the curve fitting model. The curve fitting model is based on the least-squares method to reflect the relationship between physiological variables in hypoxic training and physiological variables at the sea-level training by the least-squares approximation (1,22). We compared the estimated RBC count and Hb concentration values with the measured values (hypoxic) from 6 male and 6 female athletes. The values in the Error columns in Table 4 were calculated by taking the difference between the measured values (hypoxic) and each estimation method's estimated value.

Table 4

Comparison of RBC count and Hb concentration between the measured average values and the estimated values from curve fitting and the machine learning model.

	Measured value (sea level)	Measured value (hypoxic)	Estimated value − curve fitting	Error	Estimated value − machine learning	Error
RBC ($\times 10^{12}$/L)						
1 (male)	5.02	5.39	5.30	−0.09	5.44	**+0.05**
2 (male)	4.26	4.78	4.51	−0.27	4.82	**+0.04**
3 (male)	4.71	4.73	5.00	+0.27	4.75	**+0.02**
4 (male)	5.48	5.68	5.77	+0.09	5.68	**0**
5 (male)	5.36	5.72	5.72	**0**	5.71	−0.01
6 (male)	5.49	5.65	5.78	+0.13	5.63	**−0.02**
7 (female)	3.99	4.81	4.40	−0.41	4.87	**+0.06**
8 (female)	4.16	4.43	4.58	+0.15	4.47	**+0.04**
9 (female)	3.95	4.83	4.37	−0.46	4.86	**+0.03**
10 (female)	4.35	4.67	4.77	+0.10	4.69	**+0.02**
11 (female)	4.48	4.76	4.91	+0.15	4.77	**+0.01**
12 (female)	4.48	4.99	4.93	−0.06	4.98	**−0.01**
Hb (g/dl)						
1 (male)	15.8	17.4	17.03	−0.37	17.44	**+0.04**
2 (male)	15.87	16.65	16.92	+0.27	16.65	**0**
3 (male)	15.13	15.15	15.78	+0.63	15.18	**+0.03**
4 (male)	16.17	16.65	16.92	+0.27	16.64	**+0.01**
5 (male)	14.53	15.13	15.76	+0.63	15.13	**0**
6 (male)	16.77	17.05	17.02	−0.03	17.03	**−0.02**
7 (female)	12.23	14.37	14.38	**+0.01**	14.44	+0.07
8 (female)	11.63	12.92	13.16	+0.24	12.98	**+0.06**
9 (female)	11.47	13.3	13.45	+0.15	13.35	**+0.05**
10 (female)	12.27	13.12	13.31	+0.19	13.16	**+0.04**
11 (female)	13.17	13.57	13.66	+0.09	13.60	**+0.03**
12 (female)	13.70	13.82	13.88	+0.06	13.83	**+0.01**

Hb = hemoglobin; RBC = red blood cell.

Discussion

This article proposes a machine learning-based method to estimate the physiological variables during hypoxic training. The machine learning model presented in this study can accurately estimate the physiological variables in speed skating athletes during hypoxic training based on the athletes' physiological variables measured during training at sea level. Our results (Tables 1 and 2 and Figure 4) showed that the average RBC count and Hb concentration values of athletes increased by more than 5% after the start of hypoxic than those of training at sea level.

In data analysis, curve fitting is a commonly used method to assess the relationship between variables (22). In this study, we used curve fitting to assess the relationship between physiological variables measured at sea level and the same variables as measured during hypoxic training conditions. We obtained a polynomial function that represented the relationship between these 2 groups of physiological variable measurements. Although this method can estimate the changes of the physiological variables, it is not accurate, especially when there is a large individual difference in physiological conditions or performance levels. For example, 2 athletes showed similar RBC count and Hb concentration levels during training at sea level, but these values differed in these 2 athletes during hypoxic training. In this case, the curve fitting method is bound to have errors.

Based on the results presented in Table 4, the 2 physiological variables estimated by the machine learning model using the LSTM RNN were both closer to the measured values than curve fitting, regardless of sex. The individual difference needs to be taken into consideration when assessing the changes in physiological variables over time. Based on our results, we found that some athletes maintained a relatively stable RBC count and Hb concentration when transitioning from training at sea level to hypoxic training while some athletes showed a higher amount of change in their physiological variables. The input data of the machine learning model from the mathematical model base on physiological variable measurements. Compared with the input variables of the curve fitting model, those of the machine learning model carried more information, including day-to-day variance of RBC count and Hb concentration during training at sea level. Thus, the machine learning model using the LSTM RNN showed better accuracy.

Practical Applications

Hypoxic training typically lasts 3 to 6 weeks. One of its main functions is to improve athletes' ability in the short term before major competitions (17). It can even affect an athlete's performance in competitions if they start training in a hypoxic environment that is not suitable (3). According to the third group of data in Table 4, we found that this athlete's physiological variables did not change significantly after hypoxic training. This result indicates that the current hypoxic environment has little effect on the athlete's physiological variables. Using the machine learning method, we can estimate physiological variables during hypoxic training to avoid inefficient hypoxic training. This article proposes a method using machine learning to estimate the changes of physiological variables during hypoxic training. This study demonstrated the feasibility of the proposed method and the accuracy of the model. This method can also be used to estimate other physiological variables. By estimating athletes' physiological variables after completion of hypoxic training, coaches can know some physiological variables of athletes after hypoxic training in advance. The estimated changes in physiological variables could serve as references for coaches.

Acknowledgments

The authors give thanks to the assistance received from Jilin Institute of Physical Education, Jilin province, China, and the support received from the school of Instrument Science and Electrical Engineering, Jilin University. The research is funded by Jilin Provincial Science and Technology Department under Grant 20170204057GX and Graduate Innovation Fund of Jilin University under Grant 101832020CX184.

References

1. Chen D, Meng QH. Health status detection for patients in physiological monitoring. *Conf Proc IEEE Eng Med Biol Soc* 4: 4921–4924, 2011.
2. Dostal J, Stasek J, Kockova R, et al. Cardiac etiology of exercise induced hypoxemia within elite athletes: 3176 board #222 may 31 3:30 PM - 5:00 PM. *Med Sci Sports Exerc* 51: 3176, 2019.
3. Hamlin MJ, Hellemans J. Effect of intermittent normobaric hypoxic exposure at rest on haematological, physiological, and performance parameters in multi-sport athletes. *J Sports Sci* 25: 431–441, 2007.
4. Holliss BA, Burden R, Jones AM, et al. Eight weeks of intermittent hypoxic training improves submaximal physiological variables in highly trained runners. *J J Strength Cond Res* 28: 2195–2203, 2014.
5. Horvat T, Job J. The use of machine learning in sport outcome prediction: A review[J]. *Wiley Interdiscip Rev Data Mining Knowledge Discov* 10: 1–28, 2020.
6. Humberstone-Gough CE, Saunders PU, Bonetti DL, et al. Comparison of live high: Train low altitude and intermittent hypoxic exposure. *J Sports Ence Med* 12: 394–401, 2013.
7. Hoffmann JJML, van den Broek NMA, Curvers J. Reference intervals of extended erythrocyte and reticulocyte parameters. *Clin Chem Lab Med* 50: 941–948, 2012.
8. Hoffman HJ, Johnson RE. Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. *J Agric Biol Environ Stat* 20: 156–171, 2015.
9. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 9: 1735–1780, 1997.
10. Jacobs RA, Rasmussen P, Siebenmann C, et al. Determinants of time trial performance and maximal incremental exercise in highly trained endurance athletes. *J Appl Physiol* 111: 1422–1430, 2011.
11. Khodaee M, Grothe HL, Seyfert JH, et al. Athletes at high altitude. *Sports Health* 8: 126–132, 2016.
12. Lecun Y, Bengio Y, Hinton GE, et al. Deep learning. *Nature* 521: 436–444, 2015.
13. Link RM, Bernhard S, Brendan C, et al. Maximum-likelihood estimation of xylem vessel length distributions. *J Theor Biol* 455: 329–341, 2018.
14. Little MA, Gael V, Sohrab S, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. *GigaScience* 6: 1–6, 2017.
15. Millet GP, Brocherie F. Hypoxic training is beneficial in elite athletes. *J Med Sci Sports Exerc* 52: 515–518, 2020.
16. Matthew WH, Billaut F, Aughey RJ. Live-high train-low improves repeated time-trial and Yo-Yo IR2 performance in sub-elite team-sport athletes. *J Science Medicine Sport* 20: 190–195, 2017.
17. Park HY, Hwang HJ, Park JH, et al. The effects of altitude/hypoxic training on oxygen delivery capacity of the blood and aerobic exercise capacity in elite athletes—a meta-analysis. *J Exerc Nutr Biochem* 20: 15–22, 2016.
18. Pottgiesser T, Garvican LA, Martin DT, et al. Short-term hematological effects upon completion of a four-week simulated altitude camp. *Int J Sports Physiol Perform* 7: 79–83, 2012.
19. Powers SK, Williams JH. Exercise-induced hypoxaemia in highly trained athletes. *Sports Med* 4: 46–53, 1987.
20. Pottgiesser T, Ahlgrim C, Ruthardt S, et al. Hemoglobin mass after 21 days of conventional altitude training at 1816 m. *J Ence Med Sport* 12: 673–675, 2010.
21. Pla R, Brocherie F, Le Garrec S, et al. Effectiveness of the hypoxic exercise test to predict altitude illness and performance at moderate altitude in high-level swimmers. *Physiol Rep* 8: e14390, 2020.
22. Roy RJ. Uncertainty, sensitivity, convergence, and rounding in performing and reporting least-squares fits. *J Mol Spectrosc* 191: 223–231, 1998.

23. Sareban M, Schiefer LM, Macholz F, et al. Endurance athletes are at increased risk for early acute mountain sickness at 3450 m. *Med Ence Sports Exerc* 52: 1109–1115, 2020.
24. Siebenmann C, Robach P, Lundby C. Regulation of blood volume in lowlanders exposed to high altitude. *J J Appl Physiol* 123: 957–966, 2017.
25. Siebenmann C, Cathomen A, Hug M, et al. Hemoglobin mass and intravascular volume kinetics during and after exposure to 3454m altitude. *J Appl Physiol* 119: 1194–1201, 2015.
26. Van Thienen R, Hespel P. Enhanced muscular oxygen extraction in athletes exaggerates hypoxemia during exercise in hypoxia. *J Appl Physiol* 120: 351–361, 2016.
27. Wehrlin JP, Marti B, Hallén J. Hemoglobin mass and aerobic performance at moderate altitude in elite athletes. *Adv Exp Med Biol* 903: 357–374, 2016.
28. Wilber RL. Application of altitude/hypoxic training by elite athletes. *Med Sci Sports Exerc* 39: 1610–1624, 2007.
29. Warburton DER, Gledhill N, Quinney HA. Blood volume, aerobic power, and endurance performance: Potential ergogenic effect of volume loading. *Clin J Sport Med* 10: 59–66, 2000.
30. Wilber RL. Current trends in altitude training. *Sports Med* 31: 249–265, 2001.
31. Zhao R, Yan R, Chen Z, et al. Deep learning and its applications to machine health monitoring: A survey. *Mech Syst Signal Process* 115: 213–237, 2016.