# Predicting the Internal Load of Amateur Athlete's Based on Their External Load Using Accessible Sensors

Remco Loof
University of Twente
PO Box 217, 7500 AE Enschede
the Netherlands
r.loof@student.utwente.nl

## ABSTRACT

Injuries are major determinants of an athlete's chance at success. Training programs, therefore, try to maximize the result of the training, while minimizing the risk of an injury. This is commonly done by looking at the training load, which can be described by the external- and internal load. A good training program is based on the internal load. Unfortunately, it is hard to keep track of the internal load, whereas it is relatively easy to measure the external load. As a result, it becomes interesting to investigate to what extent the external load can predict the internal load. Research regarding the predictability of the internal load based on the external load has been performed multiple times with the use of expensive sensors which are not accessible/affordable for amateur athletes. This research extends previous research by investigating the predictability of an athlete's heart rate, based on external load measures that are collected using accessible/affordable sensors for amateur athletes. Several measures for the external load have been identified and their relevance is investigated. Using these measures, the accuracy of several machine learning algorithms is evaluated. The Ridge Regression algorithm proved to be able to predict the general trend of moderate heart rates with the use of an accelerometer, gyroscope and GPS tracker.

## Keywords

Training Load, Machine Learning, Injury Prevention, Predictive Analysis

## 1. INTRODUCTION

Research by Raysmith et al has shown that injuries are major determinants of an athlete's chance at success [16]. A proper training program can thus make a massive difference for athletes. It is however challenging to estimate which exercises will maximize the result, while not causing any injuries. Therefore, it is important to determine the intensity of an exercise. A commonly used methodology for this is the training load. The training load can be seen as the impact of an exercise on an athlete's body and can be described in two ways, the external- and internal load.

The external load of an activity is independent of the athlete and can be seen as the physical work performed by an athlete [13]. The external load is perceived as clear and straightforward and can be easily measured with sensors. Due to its convenience, it is often used to describe the amount of exercises in a training program. The internal load on the other hand is more difficult to measure and less straightforward. The internal load can be seen as how intensive an athlete perceived the activity [14].The internal load is dependent on the athlete and differs from athlete to athlete, even if the same exercise is performed. The internal load is often used as a guideline for the amount of external load an athlete can carry out. Both the external- and internal load do not have one specific measure. Table 6 summarizes some of the commonly used measures for both the external- and internal load.

The importance of training on the basis of the internal load is shown by the research of Gabbett. This research has shown that the risk of injuries increases significantly when a certain threshold of internal load is exceeded [10]. This implies that insight in the internal load is essential for drawing up proper training programs. Unfortunately, measuring the internal load of an athlete is not as straightforward as keeping track of the external load. Either more complex/expensive equipment must be used or an estimate of the internal load has to be made. As a result, it becomes interesting to investigate the possibility of predicting the internal load based on the external load. This will help in preventing injuries while using convenient equipment.

Previous research in this field was successful. These researches however often used special equipment to measure the external load. The target audience for this type of equipment are professional athletes and are not used by amateur athletes due to their high prices. The equipment could however, be helpful for amateur athletes to prevent injuries. Research by Asperti et al has shown that the amount of injuries among amateur athletes is significant [5].

This research will extend previous research by using accessible/affordable sensors, to investigate the predictability of the internal load, based on the external load. This could potentially open up the possibility for amateur athletes, to get an insight into their training load and potentially prevent injuries.

## 2. RESEARCH QUESTIONS

This research will address the following research question and sub-questions.

**RQ1**: How accurately can the internal load be predicted based on the external load using sensors that are accessible for amateur athletes?

- **RQ1.1:** Which measures can be used to measure the external- and internal load, while being accessible/affordable by amateur athletes?
- **RQ1.2:** Which measures of the external load correlate the most with the internal load?
- **RQ1.3:** Which machine learning algorithm performs the best in predicting the internal load?

# 3. RELATED WORK

Previous research has been performed regarding the predictably of the internal load based on the external load. These researches have mainly looked in the predictability of the rate of perceived exertion (RPE). The RPE is a scale measure that is used to measure the intensity of a training session [12]. Two pieces of research that have investigated the predictability of the RPE based on the external load will be provided in this section.

In 2016, the predictability of a players RPE, based on several different sets of external load measures was investigated by Carey et al [7]. During a total of 3398 training session data was collected from 45 professional Australian football players. This data concerns GPS-, heart-rate-, accelerometers-, wellness- and RPE data. The data was divided into the following six categories; running, heart rate, acceleration, player load, wellness, and derived measures. The predictability of a players RPE was investigated with the use of seven different combinations of categories. Additionally, five regression algorithms and six classification algorithms were evaluated. The best performing regression algorithm was the random forest algorithm with the use of running, heart rate and derived measures as predictors. Likewise, the best classification algorithm was the random forest algorithm with the use of the same variables. Altogether it was concluded that the data recorded by GPS tracker, heart rate monitors and accelerometers can accurately predict a player's RPE.

A similar research was held in 2017. Rossi et al investigated the predictability of the RPE based on extracted measures from a GPS tracker [17]. During 160 training sessions and 35 matches, the physical activity of 22 elite soccer players was collected. This was done using a portable GPS device that was integrated with a tri-axial accelerometer and a tri-axial magnetometer. From the captured data they were able to derive 88 external load measures. Out of these measures, 53 were selected by a feature selection process. Additionally, the RPE of the football players was gathered after each session or match. Six different classification algorithms were trained and evaluated on these datasets. The ordinal regression proved to be the best classification algorithm to describe a player's RPE. This algorithm was able to predict a player's RPE for moderate values. Unfortunately, the algorithm was not able to do this for extreme values. They argue that this discrepancy could be caused by psychological factors.

These researches show that it is possible to predict the internal load of an athlete based on the external load. This research will extend previous research in two ways. First of all, a different measure for the internal load will be used, since researches in this filed have only investigated the predictability of the RPE. Secondly, this research will focus on using sensors that are more accessible/affordable for amateur athletes. With the use of these sensors, it will be investigated if it is possible to accurately predict an athlete's internal load based on their external load. This research will potentially give new insights and open up the possibility for amateur athletes, to get an insight into their training load and thereby possibly prevent injuries.

# 4. MEASURES & SENSORS

At the beginning of this research, it was determined which measures and sensors would be used to measure the internal- and external load

## 4.1 Internal measure

The heart rate was chosen as a measure of the internal load for three reasons. First of all, no research could be found which investigated the predictability of the heart rate. Secondly, due to the limited amount of time, it would benefit the research to use the heart rate, since more data points can be collected relative to scale measures. The heart rate can be measured many times during a session, whereas scale measure can only be measured after each exercise. Finally, it is a valid, reliable and easy to use measure as depicted in table 6.

## 4.2 External measures

The external load measures were identified based on the cost and ease of use of each external load measures that are given in table 6. The relevance of each measure was considered, in addition to these factors. With these considerations in mind, it was decided to use the distance, speed, training mode, acceleration, and player load as measures for the external load. These measures can either be manually registered or measured using an accelerometer, gyroscope and GPS tracker.

# 5. METHODOLOGY

## 5.1 Data collection

In order to perform this research, a new dataset had to be generated, since no data sets were available which contained the right measures and fulfilled all additional requirements.

### *5.1.1 Used equipment*

An accelerometer, gyroscope and GPS tracker were used to measure the distance, speed, acceleration and player load of a player. These sensors are present in most modern phones. An *iPhone 8* was used with an application installed called *SensorLog* [20]. This application collects all the sensor data of the phone at a sampling rate of 100Hz. Such a high rate improves the correctness of the calculations of the measures. Lastly, a *FitBit Charge 2* was used in order to collect the heart rate during the experiment. The data collected by both the *iPhone* and *FitBit* is saved to two separate CSV-files. Which both contained a timestamp and the corresponding sensor data.

### *5.1.2 Experiment*

The data for this research was collected during two training sessions of three amateur football players. These players had no injuries or health problems. The health tracker was strapped around their left wrist and the phone was located in a running belt which was strapped around their waist. It was chosen to put the phone in a running belt instead of a sports bracelet since it would exclude unwanted measurements of arm movements from the dataset. These movements do not accurately represent the acceleration and location of a player's body. During the experiment, the different training modes of the player and their respective time stamps were written down. These training modes include the following; 0. Preparations; 1. Rest and stretch exercises; 2. Warming-up; 3. Passing and shooting exercises; 4. Practical game forms. The general outline of a training session is depicted in table 1.

**Table 1: General outline training session**

| Time | Activity |
|---|---|
| 0:00 – 0:05 | Preparations (0) |
| 0:05 – 0:15 | Warming-up (2) |
| 0:15 – 0:20 | Warming-up (2) |
| 0:20 – 0:25 | Stretching (1) |
| 0:25 – 0:35 | Position game (4) |
| 0:35 – 0:40 | Rest (1) |
| 0:40 – 0:50 | Position game (4) |
| 0:50 – 0:55 | Rest (1) |
| 0:55 – 1:10 | Shooting practice (3) |
| 1:10 – 1:15 | Rest (1) |
| 1:15 – 1:30 | Match (4) |

## 5.2 Data preparing

### *5.2.1 Combining data-sets*

Each session and player generated three data sets; one of the training modes, and two CSV-files of the sensor data and the heart rate data. These datasets are combined in the following way. First, the training modes are added to the CSV-file of the phones sensor data. This is manually done by comparing the timestamps of the sensor data and the training modes. Secondly, the CSV-file of the heart rates is combined with the CSV-file of the phones sensor data which now includes the training modes. This is a complex and time-consuming process, which is performed in *Python*. The complexity of this process is due to the different rates at which the data is recorded; the heart rate is recorded at an interval of circa ten seconds, the GPS data is recorded at an interval of circa one second and the remainder of the sensor data is recorded at a rate of 100 Hz. In *Python*, these datasets are combined by comparing the timestamps of the heart rate data and the timestamps of the collected GPS data. The downside of this choice is that the sensor and heart rate data do not match precisely, this difference is however neglectable. This process results in a single CSV-file which included the data of all three datasets, a simplified illustration of the structure of the final dataset is given in table 2.

**Table 2: Simplification of the dataset's structure**

| HR | GPS | Accelerometer/Gyroscope |
|---|---|---|
| 120 | 52.2975; 6.75553 | 0.33; 0.98; 0.22; etc. |
| | | 0.43; 0.78; 0.23; etc. |
| | 52.3001; 6.75653 | 0.32; 0.90; 0.27; etc. |
| | | 0.39; 0.91; 0.26; etc. |
| 122 | 52.2988; 6.75634 | 0.42; 0.89; 0.31; etc. |
| | | 0.34; 0.45; 0.34; etc. |
| | 52.2975; 6.75622 | 0.22; 0.76; 0.22; etc. |
| | | 0.45; 0.89; 0.12; etc. |

### *5.2.2 Outlier detection*

The combination of the datasets could include outliers. These outliers could potentially negatively impact the accuracy of the machine learning algorithms, since these measurements are used in the calculation of the measures. Therefore, it is chosen to remove outliers prior to the calculations of the measures. Outliers are detected by a program that is written in *Python*. The z-score is calculated for the sensor data of the accelerometer, gyroscope and GPS tracker. The z-score indicates how much a measurement deviates from the mean. Measurements with a z-score that is greater than three are perceived as outliers and are therefore removed from the dataset.

### *5.2.3 Calculating measures & down-sampling*

The dataset is still not ready to train various machine learning algorithms after the outliers are removed. The dataset namely only includes sensor data instead of data concerning the identified external load measures of table 6. These have to be calculated beforehand.

Furthermore, the data in the dataset is collected at different rates, which results in empty cells/unknown values for the heart rate and GPS data, this is illustrated in table 2. These cells should either be filled up or removed. This is called up-sampling and down-sampling. Up-sampling could be performed by filling the empty cells with their preceding value. Down-sampling could be performed by compressing the values to the lowest sampling rate. This would imply that the sensor data of the accelerometer, gyroscope and GPS tracker would be compressed to single values for each heart rate measurement. These values should illustrate the external load that lead to a specific heart rate. It is chosen to down-sample the datasets since this would represent the actual situation the most.

The down-sampling and calculation of the identified external load measures are performed in *Python.* This process resulted in a CSV-file, which includes instances that summarize the external load which lead to a certain heart rate. The calculations of the various identified external load measures and their down-sampling process will be discussed below.

#### *5.2.3.1 Distance*

The first identified external load measure that will be discussed is the distance. Using the sensor data of the GPS tracker it is possible to calculate the distance between two data points. The sensor data includes WGS84 coordinates of the players. The distance between each data point is calculated, using a library called *Geopy* [11].

The down-sampling of the distance is performed by summing the distances that have been calculated between two heart rate measurements. This value is equal to the distance covered by a player and is the first value that represents the external load between two heart rate measurements.

#### *5.2.3.2 Speed*

The second identified external load measure that will be discussed is the speed. The speed of the phone/user is directly calculated and collected by the phone. The phone calculates the speed based on the sensor data of the GPS tracker [2]. As a result, no additional calculations have to be made in order to obtain the speed of the player.

It is however still necessary to do some calculation for the down-sampling of the speed. The down-sampling of the speed is done by calculating the average of the speed measurements between two heart rate measurements. The average speed is the second value that represents the external load between two heart rate measurements.

#### *5.2.3.3 Acceleration*

The third identified external load measure that will be discussed is the acceleration. The actual acceleration (x, y, z) of the user/phone (taking gravity into account), is similar to the speed also directly calculated and collected by the phone. The sensor data that is used to calculate the acceleration of the user is the sensor data of the accelerometer (acceleration) and gyroscope (orientation) [3]. With this data, the magnitude of the total acceleration of each measurement can be calculated using vector calculations. This is done by using the following formula;

$$a_{total} = \sqrt{ax^2 + ay^2 + az^2}$$

The variables ax, ay, and az represent the player’s acceleration for the x, y, and z-axis.

The down-sampling of the acceleration is performed by calculating the average of all the acceleration measurement between two heart rates. The average acceleration is the third value that represents the external load between two heart rate measurements.

#### *5.2.3.4 Player load*

The fourth identified external load measure that will be discussed is the player load. This measure is calculated using the same acceleration measurements as the acceleration. The player load for a single measurement is calculated using the following formula;

$$pl = \sqrt{(ax_i - ax_{i-1})^2 + (ay_i - ay_{i-1})^2 + (az_i - az_{i-1})^2}$$

The $ax_i$, $ay_i$ and $az_i$ represent the player's acceleration for the x, y and z-axis of a measurement. The $ax_{i-1}$, $ay_{i-1}$, $az_{i-1}$ represent the player's acceleration for the x, y, z of the previous measurement [15].

The down-sampling of the player load is performed by calculating the player load per second. This is done by summing the player loads that have been calculated between two heart rate measurements. Subsequently, the sum is divided by the time interval between two heart rate measurements. The player load per second is the fourth value that represents the external load between two heart rate measurements.

#### *5.2.3.5 Training mode*

The fifth identified external load measure that will be discussed is the training mode. The training modes are already present in the dataset and do not have to be calculated or down-sampled. The training mode is the fifth value that represents the external load.

### *5.2.4 Rolling Window*

After the external load measures have been calculated and down-sampled, only one process is left to be performed. The captured data is a sequence of observations taken sequentially in time, which is called a time series. This implies that future observations are affected by previous observations, i.e. the external load that is carried out will not only influence the current heart rate but also future heart rates. This is modeled in the dataset with a rolling window. This window summarizes the external load that has been carried out prior to a heart rate measurement and shifts after each data point. A window size of six data points is used, which equals approximately half a minute of previous delivered external load. Each window contains the sum of the distance covered, and the average of the average speed, average acceleration and the player load per second of the previous six data points. For each data point, these values are calculated and added as separate columns to the dataset. The training mode is excluded from the rolling window since it is a categorical variable. The dataset can now be used to evaluate the correlation of the measures with the heart rate and train various machine learning algorithms.

## 5.3 Machine Learning

### *5.3.1 Feature importance*

The importance of each external load measure is evaluated before the machine learning algorithms are trained. This is done by calculating the Spearman's correlation coefficient for each variable. This coefficient is chosen since the relations between the heart rate and the external load measures are unclear. The coefficient ranges from -1 to 1 and gives an insight into the correlation between the heart rate and the external load measures. A positive correlation indicates that the heart rate and a measure change in the same direction. A negative correlation indicates that the heart rate and the measure change in the opposite direction. A neutral correlation indicates that there is no relation between the heart rate and the measure.

### *5.3.2 Algorithm selection*

Various machine learning algorithms will be trained after the feature importance is evaluated. Predicting the internal load based on the external load can be seen as a supervised learning problem since there is an output that maps to a certain input. Relatively speaking, these concern the heart rate and the external load measures. Regression algorithms will be evaluated, since the heart rate is a continuous variable. The *Scikit-learn* library in Python offers a variety of regression algorithms. Based on the flowchart of the *Scikit-learn* library it was chosen to train the Lasso Regression (LAR), Ridge Regression (RIR) and Support Vector Regression (SVR) algorithms [18].

- Lasso Regression (LAR): is a linear regression algorithm that can be used on datasets that have multicollinearity [1]. Multicollinearity means that two or more predictor variables are strongly correlated. This is the case in the datasets of this research.
- Ridge Regression (RAR): is a similar algorithm as Lasso Regression but uses a different type of feature selection. Lasso regression uses L1 regularization, whereas Ridge regression uses L2 regularization. L2 regularization is less thorough since the weight of an independent variable will never be zero [1]. This could prove to be better since few variables are used.
- Support Vector Regression (SVR): is an adaptation of the Support Vector Machine algorithm, which is normally applied to classification models. Support Vector Regression can, however, be applied to both linear and non-linear datasets [1]. This could prove to be valuable, since variables may have a non-linear relation with the heart rate.

### *5.3.3 Training models*

The algorithms that are chosen are trained in the following way. Each dataset is trained, validated and tested separately by each machine learning algorithm since the heart rate is dependent on the athlete. A dataset is divided in a training-, validation- and testing set using a form of nested cross-validation. Normal cross-validation cannot be used since this would result in unchronological of these sets, which would make the algorithms predict the past [4]. The training set is used to initially train an algorithm by fitting the parameters of the algorithm. Subsequently, the validation set is used to prevent overfitting by fitting the hyperparameters of an algorithm. Finally, the testing set is used to evaluate the accuracy of an algorithm. This is done by calculating the difference between the actual heart rate and the predicted heart rate by the algorithm for each instance in the testing set. The average difference of the instances is calculated, which results in the mean absolute error (MAE). A low score implies that an algorithm can accurately predict the heart rate. The calculation of the MAE can be summarized in the following formula;

$$MAE = \frac{\sum |(y_i - x_i)|}{n}$$

In this formula *n* represents the total number of instances and $y_i$ and $x_i$ represents the predicted and actual value of *i* respectively [6].

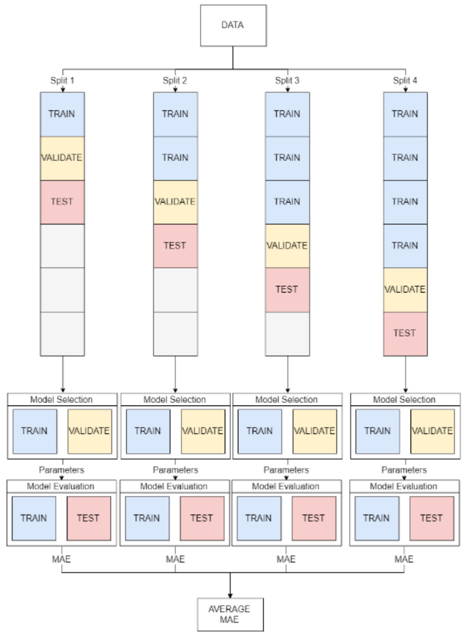


**Figure 1: Illustration of the training and evaluation process**

The splitting of the datasets is done using the day forward chaining method. Four splits are made using the datasets. Each split contains a training-, validation- and testing set and has chronologically ordered instances, in contrast to normal cross validation. This is done using the *TimeSeriesSplit* function of the *Scikit-learn* library in *Python* [19]. The overall process of the training and evaluation is illustrated in figure 1. A dataset will be divided into six sub-datasets. The first split will include the first three sub-datasets, the second split will include four sub-datasets, and so on. These sub-datasets are used to train, validate and test the algorithm. The testing of the algorithms is performed with the last sub-dataset. The training and validation of the algorithms are performed with the remainder of the sub-datasets. The overall accuracy will be calculated based on the average MAE of the four splits for each algorithm.

## 6. RESULTS

The Spearman's coefficients for each identified external load measure and participant are depicted in table 3. The average coefficient for each measure is also depicted in this table, these indicate the correlation between the heart rate and each individual measure.

**Table 3: Results feature correlation (real-time efforts)**

| Participant | Training mode | Distance | Speed | Player load | Acceleration |
|---|---|---|---|---|---|
| *1* | 0.33 | 0.077 | 0.15 | 0.25 | 0.27 |
| *2* | 0.22 | 0.24 | 0.28 | 0.39 | 0.41 |
| *3* | 0.57 | 0.5 | 0.53 | 0.58 | 0.6 |
| *Average* | *0.37* | *0.27* | *0.32* | *0.41* | *0.43* |

From the overall results of table 3, it can be concluded that each measure, a positive relationship with the heart rate has. This is logical since an increase in the external load normally results in an increase in the internal load. The player load and the acceleration have the highest correlation with the heart rate. The distance and speed, on the other hand, correlate the least with the heart rate. Interestingly, the player load and acceleration are both derived from the sensor data of the accelerometer and gyroscope, whereas the distance and speed on the other hand, are derived from the sensor data of the GPS tracker.

Unfortunately, the relationships between the external load measures and the heart rate do not seem to be as strong as expected. This is probably due to the fact that the heart rate does not increase simultaneously with the external load delivered by a player. The previous efforts of a player probably have a stronger relationship with the heart rate. The Spearman's correlation coefficients for the rolling window measures are calculated in order to test this. The Spearman's correlation coefficients for the heart rate and each rolling window measure for each participant are depicted in table 4.

**Table 4: Results feature correlation (previous efforts/rolling window)**

| Participant | Distance | Speed | Player load | Acceleration |
|---|---|---|---|---|
| *1* | 0.14 | 0.29 | 0.33 | 0.35 |
| *2* | 0.23 | 0.29 | 0.47 | 0.49 |
| *3* | 0.61 | 0.63 | 0.71 | 0.71 |
| *Average* | *0.33* | *0.40* | *0.50* | *0.52* |

Table 4 shows a slight increase in the overall correlations between the measures and the heart rate compared to table 3. The player load and acceleration still correlate the most with the heart rate, whereas the distance and speed still correlate the least with the heart rate. It can be concluded that the overall correlation of all measures is higher for the previous delivered efforts compared to the real-time efforts. However, still, neither of these measures have a strong correlation with the heart rate.

All measures were used to train and evaluate the Lasso, Ridge Regression and SVR machine learning algorithms, despite their weak relationships with the heart rate. The mean absolute errors of the trained algorithms are depicted in table 5.

**Table 5: Accuracy machine learning algorithms**

| Participant | LAR | RIR | SVR |
|---|---|---|---|
| *1* | 8.60 | 7.68 | 20.23 |
| *2* | 10.61 | 9.37 | 18.01 |
| *3* | 7.53 | 7.769 | 20.74 |
| *Average* | *8.91* | *8.27* | *19.66* |

From these results, it can be concluded that the Ridge Regression algorithm performed the best with an overall mean absolute error of 8.27. The algorithm under-/overestimated the heart rate on average by 8.27. The Lasso Regression algorithm performed similarly with a mean absolute error of 8.91. It can be concluded that the L2 regularization is a better fit for this dataset since the Ridge Regression algorithm is outperforming the Lasso Regression algorithm. The Support Vector Regression algorithm performed drastically worse compared to the other two algorithms. This is probably due to the limited size of the dataset and the limited number of features present in the dataset.

Overall, the accuracy of the algorithms appear to be mediocre, since they are not able to accurately predict the heart rate of a player. However, by taking a closer look at the actual heart rate compared to the predictions of the algorithms, it can be seen that the algorithms are able to predict the trend of the heart rate. Around 90 chronologically ordered heart rate measurements (blue line) of participant 1 are compared to the predicted heart rate of each algorithm (yellow, green and red lines), and are depicted in figure 2.

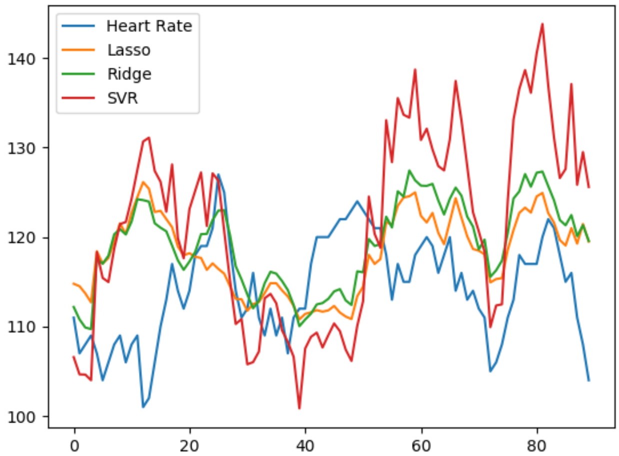


**Figure 2: Actual heart rate vs predicted heart rate**

Two things can be taken away from figure 2. The first thing that stands out concerns the predictive capabilities of the algorithms. It seems that the line of the heart rate and the lines of the predictions follow the same trend. This indicates that the algorithms can predict the trend of the heart rate in general. However, the lines of the algorithms and the line of the heart rate almost never perfectly align. The algorithms either over- or underestimated the heart rate. A clear instance of overestimation occurs between measurement 5 and 15. The internal load during this period was relatively low compared to the external load measured. An instance of underestimation occurs between measurement 40 and 50. The internal load during this period was relatively high compared to the external load measured. The second thing that stands out is the overall characteristics of the heart rate. It can be seen that the heart rate ranges from 100 to 150 and deviates quite a lot within this range. The heart rate between two successive heart rate measurements sometimes increases/decrease extremely (extreme slope) or not at all (no slope). An instance of an extreme increase occurs from measurement 17 to 18. An instance of zero increase in the heart rate occurs from measurement 41 to 42. These opposite events illustrate the unpredictability of the heart rate.

## 7. DISCUSSION

A few remarks concerning the results will be made before a conclusion will be drawn.

First of all, it has to be noted that the used datasets probably contain a significant amount of noise. The noise could explain the inaccuracy of the algorithms. The noise in the datasets could be caused by several elements. The first element that could cause noise, is the inaccuracy of the GPS tracker. A mobile GPS tracker can locate a phone within a radius of 4.9 meters [8]. The actual location of the phone can deviate from the recorded location. This implies that the distance covered between two measuring points will deviate from the actual distance covered. This could explain the low correlation between the heart rate and the measures which relied on the sensor data of the GPS tracker. The second element that causes noise in the dataset is the inconsistency of the *FitBit*. This device would sometimes stop recording the heart rate and populated the dataset with incorrect values, by registering the preceding measurement. This could be an explanation for the same heart rate measured at two successive measurements.

Secondly, it should be noted that the heart rate, a very unpredictable variable is. A player's heart rate is dependent on more factors than the external load. The results showed that the algorithms either over- or underestimation of the heart rate. Overestimation of the heart rate could occur due to a relatively high measured external load. This could happen when a player is, for example, setting out an exercise and bends over to pick up a cone or vest. An increased external load will be measured by the accelerometer, whereas this activity normally would not result in an increased heart rate. This could result in overestimation of the heart rate by the algorithms. Underestimation of the heart rate could occur when a relatively low external load is measured. An example of such event is when a player stands still after an intensive sprint. A low external load will be measured, whereas the player's heart rate will be relatively high. This could explain the underestimation of the heart rate by the algorithms. Another explanation could be the effects of psychological factors on the heart rate [9]. A player could experience pressure, which would increase the heart rate but does not increase the external load. This could lead to underestimation of the heart rate by the algorithms. Overall, it remains unclear how an increase in the external load will impact the heart rate. The correlations between the heart rate and the external load measures have shown that the relation between the external load measures is stronger for the previous delivered efforts compared to the real-time efforts. However, it remains unclear when and how exactly, the heart rate will respond to an increase/decrease of the external load.

Lastly, two remarks concerning the dataset should be made. First of all, the sizes of the datasets used in this research are limited. Circa 500 data points were provided in three different datasets. A small dataset does not necessarily explain the low accuracy of a machine learning algorithm. However, the combination of the limited size and noise in the dataset tends to explain the low accuracy. The limited size of the dataset will namely strengthen the effects of the noise on the accuracy of the algorithms. Furthermore, it should be noted that the heart rate measurements in the datasets do not cover a large amount of extreme heart rate measurements. Figure 2 has shown that the heart rates mainly range from 100 to 150. This means that the conclusion that will be drawn only applies to moderate heart rates. Unfortunately, due to time constraints and limited availability of equipment, it was not possible to gather more data.

## 8. CONCLUSION

This research investigated the possibilities of using sensors that are accessible/affordable for amateur athletes to predict the heart rate of an athlete. Five external load measures have been identified using the price and ease of use of the external load measures in table 6. The relevance of each measure was considered, in addition to these factors. The external load measures that have been identified include the training mode, distance, speed, acceleration, and player load. The sensors of a mobile phone were used to measure these measures. These sensors include an accelerometer, gyroscope and GPS tracker.

Subsequently, an experiment was set up which measured the identified external load measures and the heart rate of three players during two training sessions. The correlation of each measure with the heart rate was investigated, using the data that was collected during the experiment. Overall, the previous external load correlated more with the heart rate compared to the real-time external load. The measures that correlated the most were the player load and the acceleration, both were derived from the sensor data of the accelerometer and gyroscope. The measures that were derived from the GPS tracker correlated the least with the heart rate. These include the distance and speed.

Finally, various datasets were used to train and evaluate three machine learning algorithms. The Ridge Regression algorithm proved to be the best at predicting the heart rate based on the identified external load measures. The Ridge Regression algorithm had a mean absolute error of 8.27 The accuracy of this algorithm appears to be mediocre. However, the algorithm is able to predict the trend of moderate heart rates in general.

To conclude, this research was not able to accurately predict the heart rate of an athlete in real time, using an accelerometer, gyroscope and GPS tracker. However, this research was able to predict the general trend of moderate heart rates using these sensors. It is assumed that a combination of noise, the unpredictability of the heart rate and a limited amount of data resulted in relatively low accuracy of the algorithms compared to similar researches.

## 9. FUTURE WORKS

The results of this research could be improved by taking a different approach. This section will discuss how future data gathering should be performed in order to not only improve the accuracy of various machine learning algorithms but also to be able to draw more valuable and concrete conclusions.

First of all, it can possibly prove to be useful to set up a specific training session beforehand. This will allow the researcher to take control of the player's heart rate to some extent. This will result in an enriched dataset which contains more extreme values. By doing so it will be possible to investigate the accuracy of the machine learning algorithms for more extreme heart rate values as well. Explicitly taking control over a training session should however be done carefully. Some bias may slip in the dataset, e.g. an endurance exercise probably focusses on the distance and neglects the player load. This may result in a higher correlation of the heart rate with the distance but will also result in a lower correlation of the heart rate with the player load.

Secondly, it may prove to be useful to identify more external load measures. In this research, only five measures were identified. These measures did not strongly correlate to the heart rate, which resulted in over- and underestimations of the heart rate. The understanding of the measured external load could be improved, by identifying additional external load measures. Which should increase the accuracy of the machine learning algorithms.

Lastly, it will probably prove to be of most value to increase the size of the datasets. The datasets used in this research contain a limited amount of data. The datasets will become more reliable by collecting more data over the course of more session. An advantage of a larger dataset is that the noise in the dataset will become less apparent. This will probably increase the accuracy of the machine learning algorithms. However, more importantly, more concrete and valuable conclusions can be drawn.

## 10. REFRENCES


[1] Acharjee, Animesh & Finkers, Richard & Visser, Richard & Maliepaard, Chris. (2013). Metabolomics : Open Access Comparison of Regularized Regression Methods for ~Omics Data. Metabolomics. 3. 10.4172/2153-0769.1000126.

[2] Apple. (2019). Documentation Core Location. Retrieved from: https://developer.apple.com/documentation/corelocation/cllocation

[3] Apple. (2019). Documentation Core Motion. Retrieved from: https://developer.apple.com/documentation/coremotion/cmdevicemotion

[4] Arlot, Sylvain & Celisse, Alain. (2009). A Survey of Cross Validation Procedures for Model Selection. Statistics Surveys. 4. 10.1214/09-SS054.

[5] Asperti, Andre & Fernandes, Tiago & Pedrinelli, André & Hernandez, Arnaldo. (2017). Sports injuries among amateur athletes at a brazilian university. Acta Ortopédica Brasileira. 25. 93-98. 10.1590/1413-785220172502165651

[6] Botchkarev, Alexei. (2018). Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology.

[7] Carey, David & Ong, Kok-Leong & E. Morris, Meg & Crow, Justin & M Crossley, Kay. (2016). Predicting ratings of perceived exertion in Australian football players: Methods for live estimation. International Journal of Computer Science in Sport. 15. 63. 10.1515/ijcss-2016-0005.

[8] Diggelen, Frank, Enge, Per. (2015)The World's first GPS MOOC and Worldwide Laboratory using Smartphones. Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation

[9] Fatisson, Julien & Oswald, Victor & Lalonde, Francois. (2016). Influence diagram of physiological and environmental factors affecting heart rate variability: an extended literature overview. Heart International. 11. 10.5301/heartint.5000232.

[10] Gabbett, Tim. (2010). The Development and Application of an Injury Prediction Model for Noncontact, Soft-Tissue Injuries in Elite Collision Sport Athletes. Journal of Strength and Conditioning Research, 24(10): 2593-2603, October 2010.

[11] GeoPy. (2018). Retrieved from: https://geopy.read-the-docs.io/en/stable/#module-geopy.distance

[12] Haddad, Monoem & Stylianides, Georgios & Djaoui, Léo & A, Dellal & Chamari, Karim. (2017). Session-RPE Method for Training Load Monitoring: Validity, Ecological Usefulness, and Influencing Factors. Frontiers in Neuroscience. 11. 612. 10.3389/fnins.2017.00612.

[13] Halson, Shona. (2014). Monitoring Training Load to Understand Fatigue in Athletes. Sports medicine (Auckland, N.Z.). 44. 10.1007/s40279-014-0253-z.

[14] Impellizzeri, Franco & Marcora, Samuele & Coutts, Aaron. (2019). Internal and External Training Load: 15 Years On. International Journal of Sports Physiology and Performance. 14. 1-4. 10.1123/ijspp.2018-0935.

[15] Nicolella, Daniel & Torres-Ronda, Lorena & Saylor, Kase & Schelling, Xavi. (2018). Validity and reliability of an accelerometer-based player tracking device. PLoS ONE. 13. e0191823. 10.1371/journal.pone.0191823.

[16] Raysmith, Ben & Drew, Michael. (2016). Performance success or failure is explained by weeks lost to injury and illness in elite Australian Track and Field athletes: a 5-year prospective study. Journal of Science and Medicine in Sport. 19. 10.1016/j.jsams.2015.12.515.

[17] Rossi, Alessio & Perri, Enrico & Trecroci, Athos & Savino, Marco & Alberti, Giampietro & Marcello Iaia, Fedon. (2017). GPS Data Reflect Players' Internal Load in Soccer. 10.1109/ICDMW.2017.122.

[18] Scikit-learn. Choosing the right estimator. Retrieved from: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

[19] Scitkit-learn. Sklearn model selection.TimeSeriesSplit. Retrieved from: https://scikit-learn.org/stable/modules/ generated/sklearn.model_selection.TimeSeriesSplit.html

[20] SensorLog. (2019). Retrieved from: http://sensorlog.berndthomas.net/

**Table 6. Summary and Evaluation of Some Common Methods Used to Monitor Athlete Training Load and/or Responses**

| Method | Cost | Hardware needed | Software needed | Ease of use | Valid | Relia-ble | Used to inter-pret | Used to prescribe | Variables |
|---|---|---|---|---|---|---|---|---|---|
| **Internal Measures** | | | | | | | | | |
| RPE | L | N | Y/N | H | M-H | M-H | Y | Y | Single variable in AU (time dependent) |
| Session rating of perceived exertion | L | N | Y/N | H | M-H | M-H | Y | Y | Single variable in AU (time dependent) |
| TRIMP | L-M | Y | Y | M | M-H | M-H | Y | N | Single variable in AU (time dependent) |
| Wellness questionnaires | L | N | Y/N | M-H | M | M-H | Y | Y/N | Ratings, checklist, AU scale measures |
| Psychological inventories (e.g. POMS. Rest-Q-Sport) | L-M | N | Y/N | M-H | M-H | M-H | Y | Y | Ratings, checklist, AU scale measures |
| Heart-rate indices | L-M | Y | Y | H | H | M-H | Y | Y | Heart rate, time in zones, HR variabil-ity/recovery measures, etc. |
| Oxygen uptake | H | Y | Y | L | H | H | Y | Y | VO2, metabolic equiv-alents |
| Blood lactate | M | Y | Y/N | M | H | H | Y | Y | Concentration |
| Biochemical/hematological assessments | M-H | Y | Y/N | L | H | M-H | Y | Y | Concentrations, vol-umes |
| **External Measures** | | | | | | | | | |
| Time | L | Y | Y/N | H | H | H | Y | Y | Units of time (s, min, h, d, wk, y) |
| Training frequency | L | N | N | H | H | H | Y | Y | Session count |
| Distance/mileage | L | Y/N | Y/N | H | H | H | Y | Y | Units of distance (m, km) |
| Movement repetition counts | L | Y/N | Y/N | M-H | H | M-H | Y | Y | Activity counts (e.g. Steps, jumps, throws) |
| Training mode | L | Y?N | N | H | H | H | Y | Y | Weight training, run, cycle, swim, row, etc. |
| Power output | M-H | Y | Y | L-M | H | H | Y | Y | Relative (W/kg) and absolute power (W) |
| Speed | L-M | Y | Y/N | M-H | H | H | Y | Y | Speed measures (m/s, m/min, km/h) |
| Acceleration | L-M | Y | Y | L | H | H | Y | Y | Acceleration measures (m/s2) |
| Functional neuromuscular tests | L-M | Y | Y/N | M | M-H | H | Y | Y | Countermovement jump and drop-jump measures |
| Acute chronic-workload ratio | L-M | Y/N | Y | M | M-H | M-H | Y | Y | Size of acute training load relative to chronic load |
| GPS measures | M | Y | Y | M | M-H | M | Y | Y | Velocity, distance, ac-celeration, time in zones, location |
| Metabolic power | M | Y | Y | L-M | L-M | M | Y | N | Energy equivalent |
| Time-motion analysis video (automated) | H | Y | Y | L | M-H | M | Y | Y | Velocity, location, ac-celeration |
| Time-motion analysis video (nonautomated) | M-H | Y | Y | L | M-H | M | Y | Y | Velocity, location, ac-celeration |
| Accelerometry | M | Y | Y | L-M | M-H | M | Y | N | x-y-z g force |
| Player load | M | Y | Y | M | M | M | Y | Y | Single variable in AU (time dependent) |

## APPENDIX

## A. CONSENT FORMS

UNIVERSITEIT TWENTE.

# CONSENT FORM

**STUDY TITLE:** Predicting the Internal Load of Athletes' Based on Their External Load
**RESEARCHER:** Remco Loof

I confirm that the researcher has explained the research and the elements of informed consent to me.

I confirm that I have had the opportunity to ask questions and the researcher has answered any questions about the study to my satisfaction.

I understand that my participation is voluntary and that I am free to withdraw from the project at any time, without having to give a reason and without any consequences.

I understand that I can withdraw my data from the study at any time.

**If you have any other questions or complaints about this research you can contact me via phone: +31 06 37 32 68 82 or email: r.loof@student.utwente.nl**

*Participant Name* Sam Von Piekartz

*Participant Signature* SmV Pieb

1

UNIVERSITEIT TWENTE.

# CONSENT FORM

**STUDY TITLE:** Predicting the Internal Load of Athletes' Based on Their External Load
**RESEARCHER:** Remco Loof

I confirm that the researcher has explained the research and the elements of informed consent to me.

I confirm that I have had the opportunity to ask questions and the researcher has answered any questions about the study to my satisfaction.

I understand that my participation is voluntary and that I am free to withdraw from the project at any time, without having to give a reason and without any consequences.

I understand that I can withdraw my data from the study at any time.

**If you have any other questions or complaints about this research you can contact me via phone: +31 06 37 32 68 82 or email: r.loof@student.utwente.nl**

*Participant Name* Quincy P

*Participant Signature*

1

UNIVERSITEIT TWENTE.

# CONSENT FORM

**STUDY TITLE:** Predicting the Internal Load of Athletes' Based on Their External Load
**RESEARCHER:** Remco Loof

I confirm that the researcher has explained the research and the elements of informed consent to me.

I confirm that I have had the opportunity to ask questions and the researcher has answered any questions about the study to my satisfaction.

I understand that my participation is voluntary and that I am free to withdraw from the project at any time, without having to give a reason and without any consequences.

I understand that I can withdraw my data from the study at any time.

If you have any other questions or complaints about this research you can contact me via phone: +31 06 37 32 68 82 or email: r.loof@student.utwente.nl

*Participant Name* Lars Binnendyk

*Participant Signature* Lars B.

1