Stress Predictors and its effect on Athletic Performance

Background

Stress can be easily understood as an unspecified response of the body to any demand posed upon it. It is a substantial imbalance between psychological and physical demands placed on an individual and their response capability under conditions where failure to meet demands has important consequences.

In competitive sports stress can be a double edged sword . When used effectively, it can enable an athlete to perform at their peak and thrive under the high pressure environment of game day. On the flipside, a disbalance of stressors and recovery time can cause physiological and cognitive errors that would mean the difference between a coveted win or loss. Psychological stressors have been proven to impair muscular strength recovery after strenuous exercises [1]. Furthermore, lacking the necessary resilience to deal with a stressful situation may result in anxiety or negative thoughts that can affect performance [2].

Typically, stress can have a somatic, cognitive or behavioural impact on an athlete. Somatic effects are best described as physical effects within the human body. The stress reaction is caused by adrenaline and is popularly known as the fight or flight response. When stressed the sympathetic nervous system is activated causing an increase in adrenaline whereas when an athlete is no longer stressed the parasympathetic nervous system takes over counterbalancing the effects of the sympathetic nervous system. Cognitive stress can lead to lower reaction times [*] that have a negative effect on performance and also cause an athlete to feel the somatic effects such as breathing and sweating. The behavioural impact comes from the ability of an athlete to deal with a stressful situation. They can deal with stress either positively or negatively. Someone who fears competition would view it as a stressful situation and is most likely to not perform to their full potential. On the other hand, someone who is confident of performing well can extract pleasure from a competitive situation and would most probably bag a victory for their teams or themselves. This can be best illustrated if we were to imagine a world cup scenario wherein a star player who is confident of their competency would put themselves forward to take a penalty kick in the world cup final as they would get a ‘buzz’ from it whereas lesser confident players may fumble and would want to avoid such a situation. The psychological demands exceed the physical capabilities of some and therefore they are less likely to be as successful as the more confident player. The consequences however remain the same for both the players - winning or losing the world cup title.

Objective

A stress detection or prediction system can help an athlete monitor their stress and therefore control its detrimental effects. While we start with clinically accepted parameters to strengthen this predictor we can also hope to uncover possible parameters that may help in indicating someone's stress. This prediction system can go in as a feature into either the consumer app, coach app or zheal app depending on how it aligns with the product roadmap and goals.

Another secondary objective that we can hope to achieve through this module are possible hidden insights based on stress and an athlete's sleep behaviour.

Value Proposition (Productising Stress Prediction)

From a value proposition perspective the closed loop on stress for athletes can be recommending how intense their training (physical exercise) should be on the next day based on their stress levels measured during sleep. Another value add can be their risk of injury based on the physical stress their body may be in during training and their daily activities during the day. The challenge with the latter approach is that currently we do not have any data collection being done around which athletes have undergone injury therefore will first go down the route of predicting stress based on features we already have and then correlate that with injury or risk of injury.

Exploratory Data Analysis

Here we visualise probable correlations between different parameters. The correlations have been done keeping in mind the trade offs and assumptions needed between pearson, spearman and kendall's correlations [3]. These would later be used as parameters that help us predict an user's stress. For the examples below we have used samples from the replica database (revised on 8th September) and chosen to plot the values of device ids which had the maximum data points available. Data points chosen initially are based on clinically proven measures that activate the parasympathetic and sympathetic nervous systems [4] ***(Lf, Hf, Vlf, LfByHf)*** and a set of parameters ***(DeepSleep duration, LightSleep duration, RemSleep duration, WakeSleep duration, Score, Recovery, AvgSDNN, AvgStress, AvgHeartRate, AvgBreathRate)*** hypothesised to be correlated to stress in terms of duration [5] or other signals related to heart rate variability and respiration. The recovery score was taken as a feature because we would like to understand if stress can be further used to predict the ability of a person to recover as well. We intend to fine tune these set of parameters which would act as predictors of stress through experimentation :

- sleep_score:
    - Deep stage duration
    - Light stage duration
    - Rem stage duration
    - Wake stage duration
    - Score
    - Recovery
    - AvgSDNN
    - AvgStress
    - AvgHeartRate
    - AvgBreathRate

- Sleep_raw_data (yet to be tested after we have access to this table):
  - Lf
  - Hf
  - Vlf
  - LfbyHf
  - Rmssd

Dozee Data

Based on parameters collected from **sleep_scores** table :

The plots below help visualise parameters estimated to be correlated with stress and have been used to measure stress during sleep. Through the plots below we see that the following parameters are correlated to stress :

- sleep_score:
  - **Deep stage duration** (based on fig 2)
  - Light stage duration
  - Rem stage duration
  - Wake stage duration
  - Score
  - **Recovery**
  - **AvgSDNN**
  - AvgStress (target variable)
  - **AvgHeartRate**
  - **AvgBreathRate**

	DeepSleep	LightSleep	RemSleep	WakeSleep	Score	Recovery	AvgSDNN	AvgStress	AvgHeartRate	AvgBreathRate
DeepSleep	1.000000	-0.346656	-0.520943	0.129967	-0.225949	-0.288599	0.193809	-0.221012	-0.165027	0.042620
LightSleep	-0.346656	1.000000	-0.128196	-0.455814	-0.062846	-0.047291	-0.215082	0.192496	0.091960	-0.116573
RemSleep	-0.520943	-0.128196	1.000000	-0.624136	0.658106	0.494914	-0.078092	0.088886	0.041389	-0.045816
WakeSleep	0.129967	-0.455814	-0.624136	1.000000	-0.520499	-0.302002	0.124233	-0.096839	-0.000292	0.118276
Score	-0.225949	-0.062846	0.658106	-0.520499	1.000000	0.344547	0.018608	-0.003186	0.047628	-0.084781
Recovery	-0.288599	-0.047291	0.494914	-0.302002	0.344547	1.000000	0.577200	-0.586713	-0.468140	-0.353872
AvgSDNN	0.193809	-0.215082	-0.078092	0.124233	0.018608	0.577200	1.000000	-0.978466	-0.435862	-0.254168
AvgStress	-0.221012	0.192496	0.088886	-0.096839	-0.003186	-0.586713	-0.978466	1.000000	0.494199	0.289989
AvgHeartRate	-0.165027	0.091960	0.041389	-0.000292	0.047628	-0.468140	-0.435862	0.494199	1.000000	0.301294
AvgBreathRate	0.042620	-0.116573	-0.045816	0.118276	-0.084781	-0.353872	-0.254168	0.289989	0.301294	1.000000

Fig 1. Correlation plot based on Pearson's correlation on the replica dataset (revised on 8th sep)

	SHS	SD	HR	BR	Recovery	ds_time_in_mins	ls_time_in_mins	rem_time_in_mins	wake_time_in_mins	AverageSDNN	AverageStress
SHS	1.00	0.54	0.09	-0.10	0.44	0.41	0.60	0.61	-0.14	-0.04	0.07
SD	0.54	1.00	0.07	-0.18	0.68	0.66	0.94	0.89	0.08	0.02	-0.02
HR	0.09	0.07	1.00	0.17	-0.36	-0.20	0.10	0.13	0.06	-0.49	0.57
BR	-0.10	-0.18	0.17	1.00	-0.29	-0.18	-0.19	-0.13	0.06	-0.14	0.18
Recovery	0.44	0.68	-0.36	-0.29	1.00	0.65	0.65	0.64	0.02	0.58	-0.57
ds_time_in_mins	0.41	0.66	-0.20	-0.18	0.65	1.00	0.61	0.57	-0.09	0.30	-0.33
ls_time_in_mins	0.60	0.94	0.10	-0.19	0.65	0.61	1.00	0.86	0.03	-0.02	0.03
rem_time_in_mins	0.61	0.89	0.13	-0.13	0.64	0.57	0.86	1.00	0.11	-0.03	0.06
wake_time_in_mins	-0.14	0.08	0.06	0.06	0.02	-0.09	0.03	0.11	1.00	0.06	-0.07
AverageSDNN	-0.04	0.02	-0.49	-0.14	0.58	0.30	-0.02	-0.03	0.06	1.00	-0.97
AverageStress	0.07	-0.02	0.57	0.18	-0.57	-0.33	0.03	0.06	-0.07	-0.97	1.00

Fig 2. Pearson's correlation done on the latest dataset (general users)

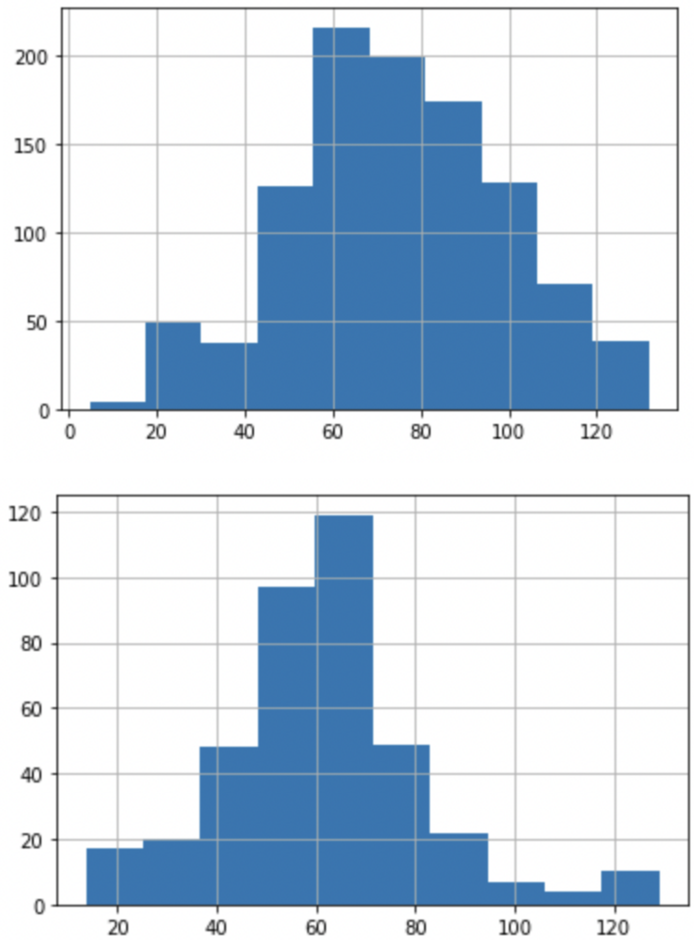


Fig 3. Distribution of Stress for general users vs athlete

date	SHS	SD	HR	BR	Recovery	ds_time_in_mins	ls_time_in_mins	rem_time_in_mins	wake_time_in_mins	AverageSDNN	AverageStress
2022-08-28	0.633068	6.566667	61.23	14.21	392.0	44.5	119.5	23.5	429.0	40.08	98.65
2022-08-28	0.790148	9.416667	56.07	14.98	2086.0	68.0	328.0	115.5	490.0	69.46	50.77
2022-08-28	0.649244	5.816667	53.68	21.13	892.0	59.0	198.5	77.5	275.0	63.95	61.28
2022-08-28	0.761281	8.616667	56.26	15.92	1731.0	77.5	261.0	155.0	515.0	64.09	60.00
2022-08-28	0.558497	6.033333	40.15	14.24	1012.0	79.0	209.5	58.0	482.0	53.91	57.24
2022-08-28	0.678005	6.350000	63.60	12.98	1272.0	65.0	214.0	101.0	518.0	56.59	71.26
2022-08-28	0.789416	8.666667	56.69	14.36	1620.0	71.5	253.5	128.0	392.0	74.07	48.25

Fig 4. An example of the data points considered in this paper captured as a snapshot of the data being used.

Given below are some visual depictions to illustrate how strongly correlated certain signals are. This should also prove why some features even though they had a seemingly low correlation can be considered to predict stress.

Furthermore, to test this, line plots were used for validation (this user - dev_ref_id = 7751, had the highest number of data points collected) :

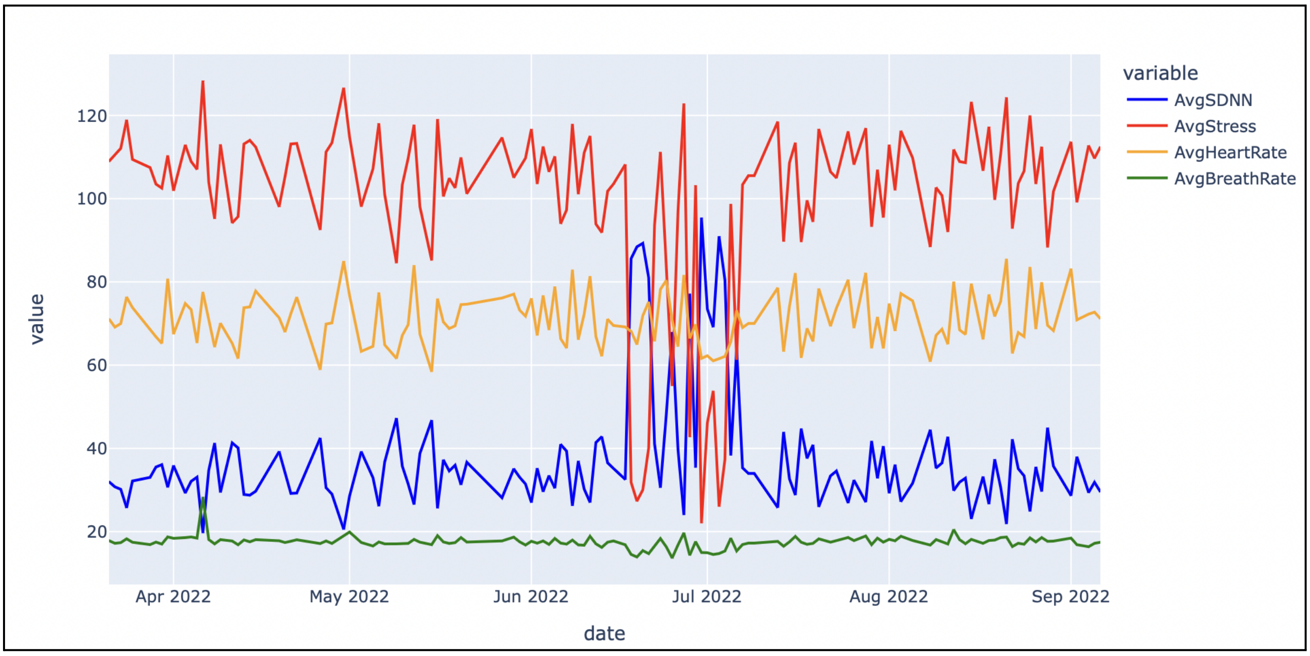


Fig 5. Correlations between AvgStress and other variables

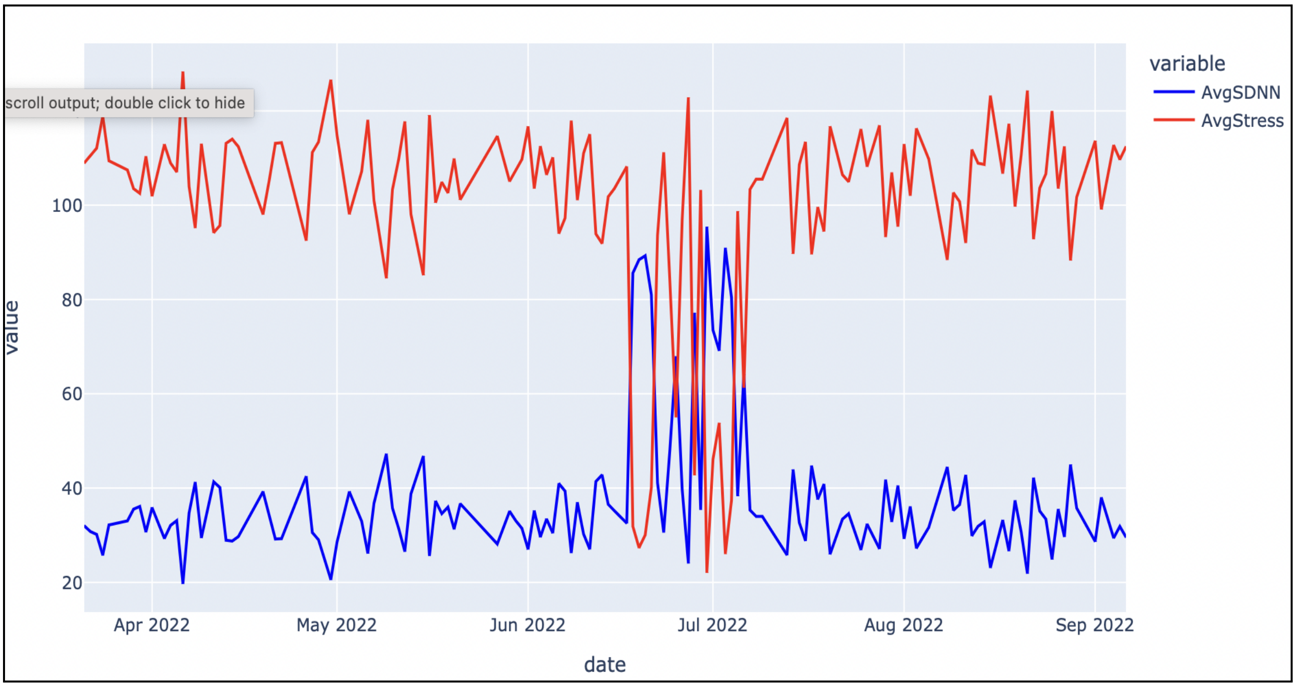


Fig 6. AvgStress correlation with AvgSDNN for one user

- AvgStress is strongly negatively correlated to AvgSDNN (higher stress->lower SDNN)

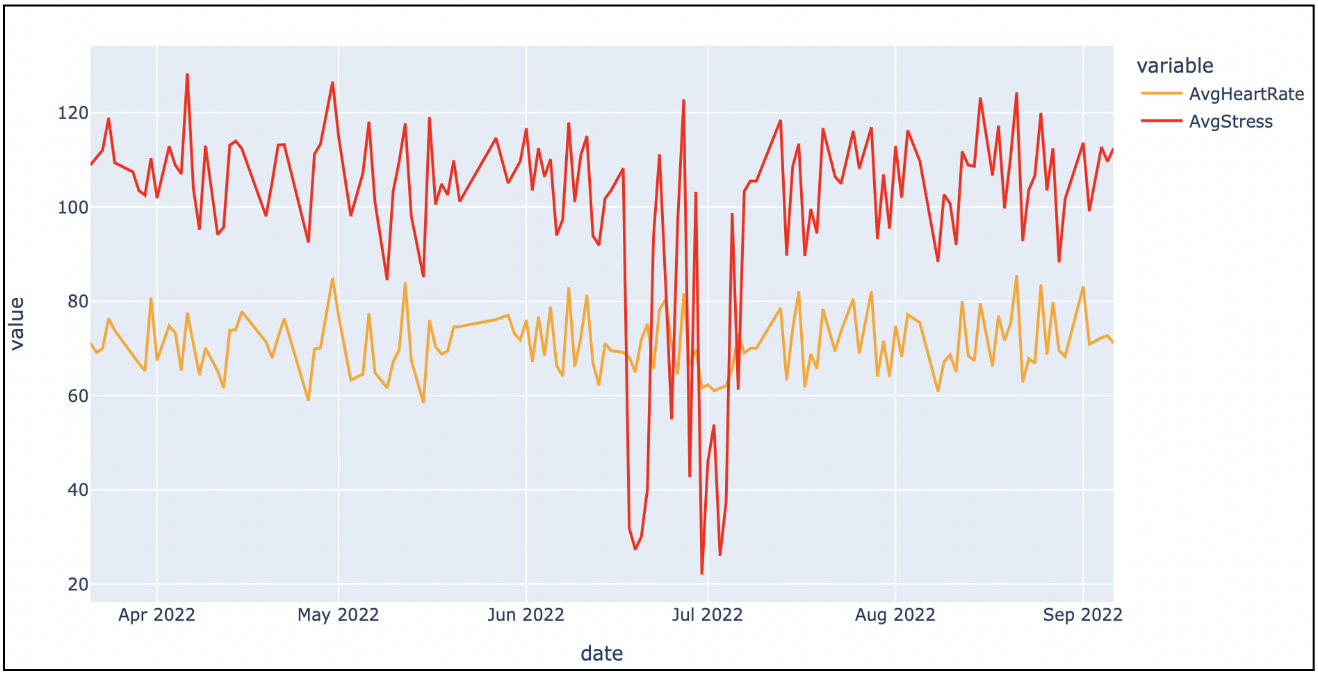


Fig 7. AvgStress correlation with AvgHeartRate for one user

- AvgStress is positively correlated to avg heart rate.

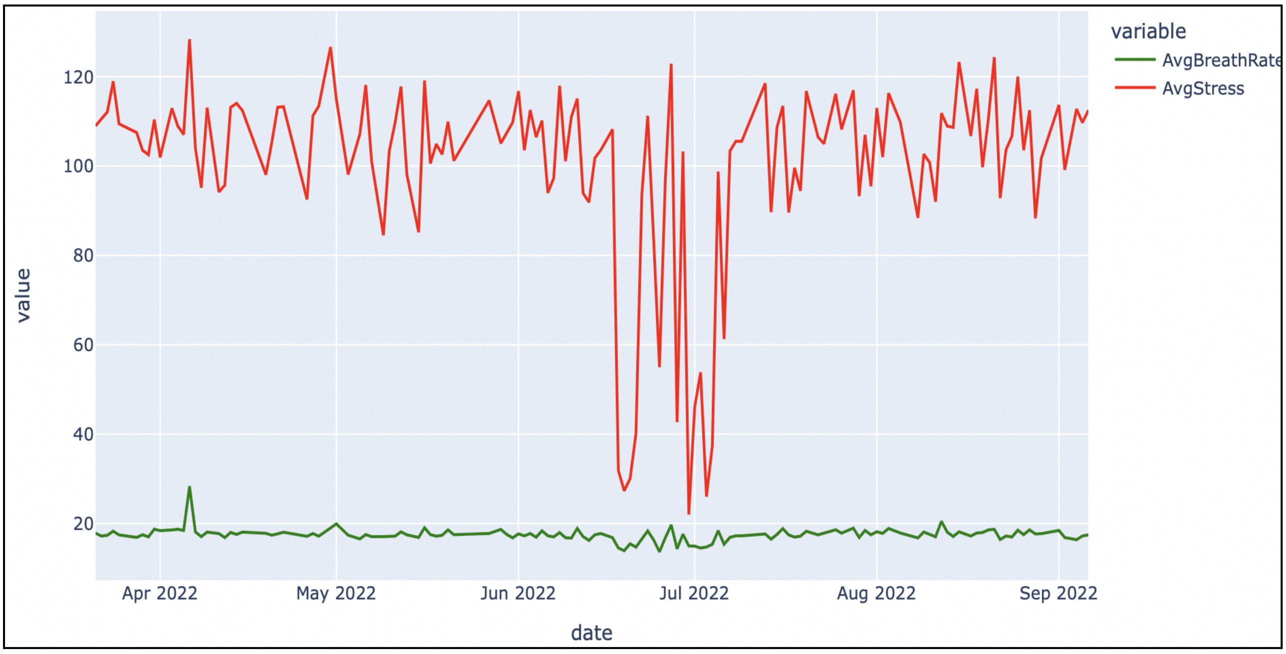


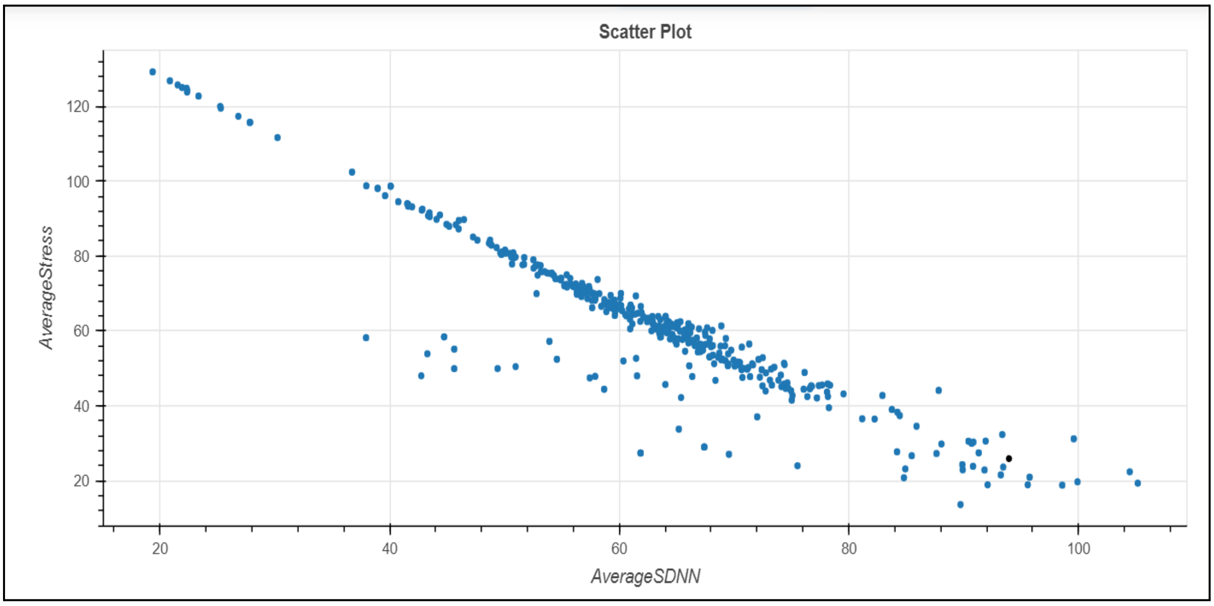
Fig 8. AvgStress correlation with AvgBreathRate for one user

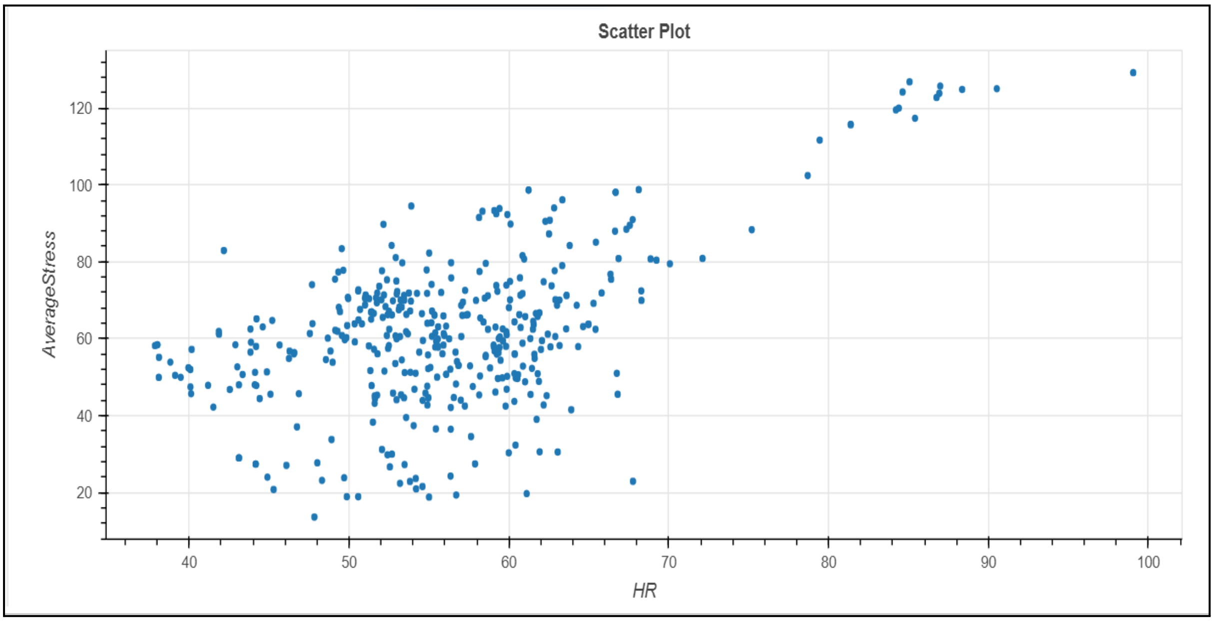
- AvgStress is faintly correlated with breath rate.

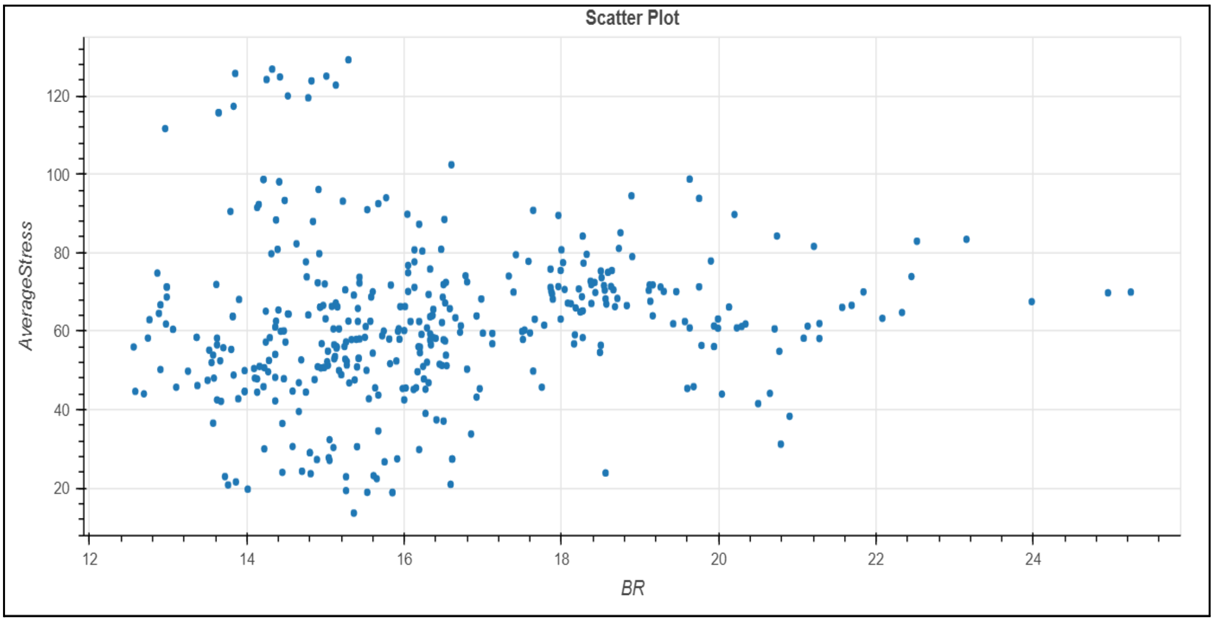
Predicting Stress

Based on the nature of the target variable that is whether they are numeric values or categorical values we can do either a regression or classification. In open source datasets like SWELL [6] where there are labels for stress we would go with a classifier. However, since we are using a device that has numeric values we tried the regression route. This decision can be later settled while optimising the model and depending on the nature of the open datasets available to us.

The scatter plots below, show the data distribution visually and the relationship between the features and average stress and therefore enable us to decide what kind of regression function would be suitable here.







AverageSDNN seems to have a very strong almost linear correlation with stress and therefore would be taken as the primary feature for regression.

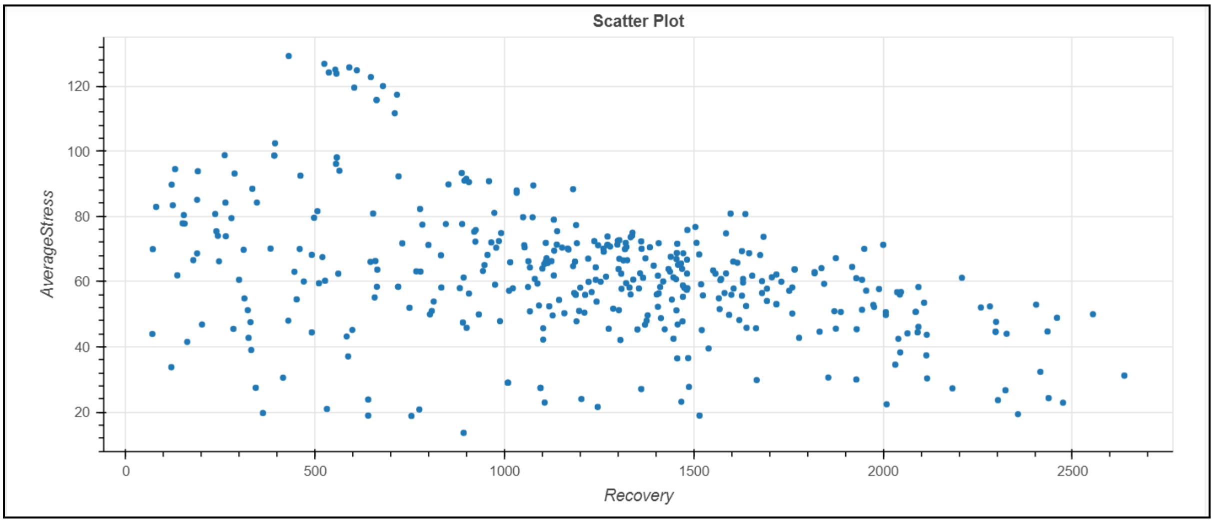
Regression based approach

Using any kind of a regression function we can hope to approach this problem. Linear regression can surely be done based on the scatter plots as shown above. Apart from that we also tried a hierarchical boosting regressor assuming some sort of non linearity amongst the features as well. While the initial set of features were chosen with very strong correlations we tried a decision tree regressor to test if we extended this set of features and also included seemingly non linear features how that would pan out for us.

Model	Features	Coefficient of determination of the prediction
Linear Regressor	AvgSDNN, AvgHeartRate, AvgBreathRate, ds_time_in_mins	**0.92056**
Hierarchical Boosting Reg	AvgSDNN, AVGHeartRate, BreathRate, ds_time_in_mins	**0.76133**

What after being able to Predict Stress ?

After predicting stress we can start to also predict how well or how soon an athlete can recover. Furthermore, using stress we can also make a correlation with recovery. We were able to get a reasonably good coefficient of determination of prediction score which indicates our ability to predict recovery using a hierarchical boosting regressor and features such as AverageSDNN, AvgBreathRate, AvgHeartRate. After adding a few more features this score was further improved to **0.69**.



Model	Features	Coefficient of determination of the prediction
Linear Regressor	AvgSDNN, AvgHeartRate, AvgBreathrate, ds_time_in_mins, AverageStress	**0.62**
Hierarchical Boosting Reg	AvgSDNN, AvgHeartRate, AvgBreathrate, ds_time_in_mins, AverageStress	**0.69**

Reference

1. Psychological stress impairs short-term muscular recovery from resistance exercise - https://pubmed.ncbi.nlm.nih.gov/22688829/
2. The Effect of Cognitive Anxiety on Sport Performances among Football Players - https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.684.937&rep=rep1&type=pdf
3. Comparison of correlation coefficients and when to choose which - https://ademos.people.uic.edu/Chapter22.html#4_comparing_correlations
4. Modulation of the Sympatho-Vagal Balance during Sleep: Frequency Domain Study of Heart Rate Variability and Respiration - https://pubmed.ncbi.nlm.nih.gov/22416233/
5. Association between Sleep Duration and Perceived Stress: Salaried Worker in Circumstances of High Workload - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5923838/
6. SWELL dataset - https://www.kaggle.com/datasets/qiriro/swell-heart-rate-variability-hrv
7. [*] The impact of Stress on body function - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5579396/pdf/EXCLI-16-1057.pdf
8. Heart Rate Variability (HRV) as a Tool for Diagnostic and Monitoring Performance in Sport and Physical Activities - https://www.researchgate.net/publication/285968540_Heart_Rate_Variability_HRV_as_a_Tool_for_Diagnostic_and_Monitoring_Performance_in_Sport_and_Physical_Activities
9. Use of heart rate variability in monitoring stress and recovery in judo athletes - https://pubmed.ncbi.nlm.nih.gov/24276307/