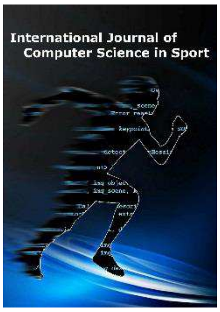
Pacing Patterns of Half-Marathon Runners: An analysis of ten years of results from Gothenburg Half Marathon

Citation for the original published paper (version of record):
Johansson, M., Atterfors, J., Lamm, J. (2023). Pacing Patterns of Half-Marathon Runners: An analysis of ten years of results from Gothenburg
Half Marathon. International Journal of Computer Science in Sport, 22(1): 124-138.
http://dx.doi.org/10.2478/ijcss-2023-0014

International Journal of Computer Science in Sport



Volume 22, Issue 1, 2023

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2023-0014

Pacing Patterns of Half-Marathon Runners: An analysis of ten years of results from Gothenburg Half Marathon

Moa Johansson, Johan Atterfors and Johan Lamm

Chalmers University of Technology, Dept. of Computer Science and Engineering, Gothenburg, Sweden

Contact: moa.johansson@chalmers.se

Abstract

The Gothenburg Half Marathon is one of the world's largest half marathon races with over 40 000 participants each year. In order to reduce the number of runners risking over-straining, injury, or collapse, we would like to provide runners with advice to appropriately plan their pacing. Many participants are older or without extensive training experience and may particularly benefit from such pacing assistance. Our aim is to provide this with the help of machine learning. We first analyze a large publicly available dataset of results from the years 2010 – 2019 (n = 423 496) to identify pacing patterns related to age, sex, ability, and temperature of the race day. These features are then used to train machine learning models for predicting runner's finish time and to identify which runners are at risk of making severe pacing errors and which ones seem set to pace well. We find that prediction of finish time improves over the current baseline, while identification of pacing patterns correctly identifies over 70% of runners at risk of severe slowdowns, albeit with many false positives.

KEYWORDS: HALF-MARATHON, RUNNING, PACING PATTERNS, RESULTS DATA, MACHINE LEARNING

Introduction

The Gothenburg Half Marathon is one of the world's largest half-marathons. Most participants are recreational runners of all ages and fitness levels, and many return to participate each year. By supporting runners to pace well, fewer runners may have to abandon the race due to fatigue, injuries or in extreme cases even collapse. A well-paced race will likely be a more pleasant experience which encourages continued running and return participation and contributes to public health (Lee et al., 2017).

We base our work on ten years of public results data (2010 – 2019) from the Gothenburg Half Marathon (n = 423 496) where finish times and 5 km split times are recorded. To our knowledge this is the largest investigation of pacing patterns for half-marathon running. Our goal is to investigate if we can use this large, easily accessible, public database to analyze and predict what is indicative of both good and bad pacing performance for recreational runners. Furthermore, it allows us to compare results on the same course with temperatures ranging between 13-25°C. As an immediate outcome, this will enable the Gothenburg Half Marathon race organizers to inform participants of risk factors related to pacing. In the longer run this work can be used in the development of tools for aiding pacing. For example, personalized pacing apps could help participants find a pace suitable for their fitness level and the conditions on the day of the race.

We formulate two connected research questions which we will aim to answer in this paper:

Research Question 1 (RQ1): Can we **(a)** use machine learning to predict half-marathon runners' finishing time? Furthermore, can we **(b)** predict which runners are at high and low risk of experience a severe slowdown during the second half of the race?

For machine learning models to even have a chance of working effectively, there needs to be statistical patterns in the data, *features*, from which the model can learn. This leads us to the second research question:

Research Question 2 (RQ2): What features from our dataset affect pacing patters and should be provided as inputs to the machine learning models?

To answer the first question, we first need to define what constitutes good and bad pacing. Optimal pacing depends on many factors (Roelands, de Koning, Foster, Hettinga, & Meeusen, 2013). As our focus is on recreational runners, the aim is not necessarily to encourage people to run as fast as possible, but rather to pace in a way to promote finishing the race in a safe manner, with low risk of overexertion or injury. In short, we optimize for long-term health benefits rather than finish time. Half-marathon runners are commonly advised to run an even or negative split: to maintain a controlled pace during the first half, and if possible then increase the pace during the second half. Whether or not an even or negative split really is the optimal pacing strategy for a half-marathon is however somewhat unclear (Abbiss & Laursen, 2008). Some earlier studies suggests that many runners in fact slow down throughout but also that half-marathon runners pace more evenly than full marathon runners (Nikolaidis, Cúk, and Knechtle 2019).

For our purposes it is sufficient to identify approximate thresholds for good and bad pacing. We adapt a definition for *severe pacing errors*, originally developed for the full marathon distance by Smyth (2021), as a severe drop in pace during a 5 km segment. Conversely, we assume that runners managing an even or negative split are unlikely to have overexerted themselves and use this as a sufficient approximation of good pacing.

To answer the second question, we need to establish which characteristics of runners could be indicative of different pacing patterns and risks, and thus useful for machine learning methods. The Gothenburg Half Marathon has been the subject of several previous studies which provides a starting point: Knechtle and Nikolaidis (2018) investigated age differences in finishing times on Gothenburg Half Marathon between 2014-2016 and found the relatively fastest finishing times for female runners aged below 40, and males between 35-39. Other studies have investigated the incidence and characteristics of runners collapsing, requiring medical assistance or ambulance transport, showing a higher incidence in warm years, and among runners younger than the average age (Carlström et al., 2019; Khorram-Manesh et al., 2020; Lüning, Mangelus, Carlström, Nilson, & Börjesson, 2019). For half-marathons in South Africa, older female runners were found to be less likely to finish races and females over the age of 50 were at higher risk of medical complications (Schwabe, Schwellnus, Derman, Swanevelder, & Jordaan, 2014a, 2014b). In an analysis of the Vienna half-marathon in 2017, results pointed toward younger and male runners being more at risk of slowdowns, while female and older runners generally paced more evenly. (Cúk, Nikolaidis, & Knechtle, 2020; Cúk, Nikolaidis, Markovic, & Knechtle, 2019). Similar patterns have been reported for the full marathon distance (Berndsen, Lawlor, & Smyth, 2020; Deaner, Carter, Joyner, & Hunter, 2015; March, Vanderburgh, Titlebaum, & Hoops, 2011; Smyth, 2021). On the full marathon distance, Ely, Cheuvront, Roberts, and Montain (2007) investigated the impact of weather and temperature and found trends towards slowing with increased wet-bulb globe temperature. Trubee, Vanderburgh, Diestelkamp, and Jackson (2014) found that for non-elite full marathon runners, female runners pace better than male, and that this was magnified in peak temperature.

As our dataset is larger than the above-mentioned studies of various half-marathons, we first conduct an exploratory data analysis to validate if the expected patterns and features from previous work is supported by our data. We investigate pacing patterns based on sex, age groups and different fitness levels (we use finish time as a proxy), as well as the effect of temperature.

Methods

We first present our dataset, followed by the metrics and definitions used to categorize good and bad pacing patterns. We then briefly summarize the statistical analysis applied to the dataset and introduce the machine learning models trained using this information as features.

Data

Our data consists of results from Gothenburg Half Marathon from the years 2010 – 2019 (earlier years did not have split times available). This data is publicly available from the race organizers website[1], we work with a snapshot of the underlying results database retrieved on 2 November 2021. Each runner is identified by a unique numeric ID and relevant to our analysis are finish time, split times at 5, 10, 15 and 20 km, year of birth and sex. In addition, we added information about the measured temperature on the race day each year, obtained from the Swedish Meteorological and Hydrological Institute. Runners start in different groups throughout the afternoon, but we simply used the temperature measured at 3pm. Note that the average daytime top temperature for Gothenburg in the month of May (when the race is held) is 17°C.

After pre-processing to remove entries with missing or obviously faulty information (e.g., missing/incorrect split- and finishing times) we obtained a dataset of 423 496 records (female = 140 409; male = 283 087). The dataset contains 184 890 unique individuals, on average participating 2.3 times in the ten-year period.

[1] https://reg.goteborgsvarvet.se/sok/resultatlista.aspx

Pacing Metrics

To identify and compare pacing patterns we use the metrics defined below:

Split Difference. The *split difference (SD)* captures time gained or lost during the second half of the race. A half marathon is 21 097.5 meters (Gothenburg Half Marathon has been measured exactly by World Athletics). Thus, as no exact mid-point split is available, we introduce a corresponding constant[2] by which we multiply the 10 km split and define SD as:

$$SD = FinishTime - 10kmSplit * 2.10975$$

$SD < 0$ indicate that the runner was faster on the second half (a negative split), while $SD > 0$ indicate they slowed down (a positive split).

Severe Pacing Error. We slightly adapt the operational definition for the full marathon distance by Smyth (2021), as a 25% slowdown on a segment, compared to an initial base-pace. Note that what we call a *severe pacing error (SPE),* Smyth refers to as “hitting the wall” (HTW).

We denote the pace of segments between split times as *pace(5 km)* for the pace of the 0-5km segment, *pace(10 km)* for the pace of the 5-10 km segment, etc. We first define the *base-pace (BP)* as the average pace over the 5 and 10 km splits. Here, the runner establishes their pacing, and the risk of severe slowdowns this early in the race is low.

$$BP = \frac{pace(5\ km) + pace\ (10\ km)}{2}$$

The BP is then used to compute the *Degree of Slowdown* (DoS) for the segments between split times in the second half of the race defined as the ratio of segment pace and base pace. For each segment s in $Segs = \{10 - 15 \text{ km}, 15 - 20 \text{ km}, 20 - 21 \text{ km}\}$, the degree of slowdown is thus:

$$DoS(s) = \frac{pace(s) - BP}{BP} = \frac{pace(s)}{BP} - 1$$

Finally, we define a SPE on a segment $s \in Segs$ as: $SPE(s) = DoS(s) \geq 0.25$. A runner has thus made a sever pacing error if for some segment $s \in Segs$, *SPE(s) = True*.

Successful Pacing. In contrast to runners making severe pacing errors, we define successful pacing as runners managing a negative or even split, as these can be assumed to be at lower risk of overexertion. As we compute SPE’s only after the 10 km mark, we do the corresponding for successful pacing and denote by *pace(start – 10km)* the pace during the up until the 10km split and *pace(10km – finish)* the pace from there to the finish, and the define the *split ratio:*

$$SplitRatio = \frac{pace(start - 10\ km)}{pace(10\ km - finish)}$$

A split ratio of ≤ 1 means the runner has managed a negative (or even) split. Note that the distance in the denominator is slightly longer, as there is no split time at the exact mid-point of the race.

[2] 1 / (10 000 / 21 0975.5) = 2.10975

Statistical Analysis

For analysis of the different features affecting the risk of making sever pacing errors, we use Python and the scipy.stats library. Our code is available online[3].

To make comparisons between pairs of groups (e.g. if male/female runners are more likely to make a SPE), we use a Fisher Exact test, provided by the Python library function scipy.stats.fisher_exact[4], which takes a 2×2 contingency table as input, and returns the resulting p-value and Odds Ratio (OR). For comparisons between multiple groups (e.g. age groups) we use a chi-square test, provided by the Python scipy.stats.chi2_contingency[5], which similarly takes a $n \times 2$ contingency table as input and provides a resulting p-value. ORs are then computed between consecutive pairs of groups. Finally, for comparisons of the effect of temperature on finishing times and proportion of runners making severe pacing errors we use a standard linear regression provided by the Python library function scipy.stats.linregress[6]

Machine Learning Models

We consider two machine learning problems. First, a regression problem of predicting the finish times of runners at different points in the race. Second, a classification problem predicting which runners will make severe pacing errors, or conversely, run a negative split. The source code is written in Python using open-source libraries and is available online[7].

Finish Time Prediction. Predicting the finishing time will be easier the further the runner has progressed through the race. We thus expect more accurate predictions when the model has access to more information. Therefore, for each method, four separate models were trained to make predictions, simulating how the runner progress through the race with the following inputs: **a)** pace at 5km split **b)** pace at 5 and 10km splits **c)** pace at 5, 10 and 15 km splits, and **d)** pace for all splits, including the one at 20km. Three machine learning methods for predicting finish time are compared:

1) A baseline model currently used for Gothenburg Half Marathon live results which simply predicts that runners will maintain their most recent 5 km pace for the remainder of the race.

2) Linear regression from the sklearn.linear_model Python library[8] (Buitinck et al., 2013). As we want to compare the accuracy at different points in the race, we train one linear regression model for each split, i.e., four in total as described above.

3) A small feed forward neural network model consisting of one hidden layer with 40 nodes, implemented using the Tensorflow library (Abadi et al., 2016). This model can capture non-linear relationships between its inputs and the finish time if such relationships be present. As with the linear regression model, we train four variants each predicting finish time at different points in the race as it progresses. We then also experiment with including additional features such as age, sex, daytime temperature, and prior finishing time for repeat participants.

[3] https://github.com/atjaoan/PacingProject/tree/main/PythonNotebooks

[4] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.fisher_exact.html

[5] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html

[6] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html

[7] https://github.com/atjaoan/PacingProject/tree/main/MachineLearningofPacingPatternsforHalfMarathon

[8] https://scikit-learn.org/stable/modules/linear_model.html

Pace Category Prediction. We also investigate if it is possible to predict which runners will make SPEs before they do so, and conversely which runners seem on track for a negative split. We phrase this as a classification problem with three classes: *SPE*, *Neg Split* and those in between, labelled *Other*. These classes are not balanced: there are roughly the same number of runners making SPEs as running negative splits, but the majority is in between. This poses a challenge as there will be less training data in the SPE and Neg Split categories. Therefore, after some preliminary experiments with standard machine learning models, we opted for a balanced Random Forest model (Lemaître, Nogueira and Aridas, 2017), which is designed for imbalanced datasets such as this. We use the implementation from the imbalanced-learn Python library[9].

Results

Pacing Patterns

Table 1 summarizes the data year by year. Overall, 9.8% of participants ran a negative or equal split, while 8.6% of runners experienced a SPE on some segment, most commonly between 15-20 km. As expected, this is a smaller proportion than in studies on the full marathon distance, where there is an increased prevalence of slowdowns due to glycogen depletion. For half-marathons, slowdowns are instead more likely due to lactate buildup or simply fatigue from overexertion during the first half.

Table 1. Summary of the data by year, number of runners, percentage of female runners, average finishing times and percentage of runners having experienced a SPE or run a negative split respectively. The warmest year was 2013 (25°) and the coldest 2012 (13,6°).

Year	Runners	%Female	Temp °C	Average time		% SPE		% Neg Split	
				M	F	M	F	M	F
2010	37 982	29.0	21.7	02:03:59 ± 00:19:35	02:15:45 ± 00:19:22	16.7	5.7	5.3	6.7
2011	42 838	30.8	16.6	01:57:06 ± 00:18:27	02:09:59 ± 00:18:46	6.2	3.8	14.6	12.1
2012	42 838	31.2	13.6	01:56:04 ± 00:19:01	02:09:10 ± 00:19:05	7.1	4.4	15.1	12.5
2013	44 919	33.0	25.0	02:05:22 ± 00:19:53	02:16:46 ± 00:20:00	16.8	7.6	6.0	8.0
2014	47 187	34.6	18.9	01:59:38 ± 00:20:24	02:13:10 ± 00:20:18	12.4	7.2	6.8	6.5
2015	46 207	34.8	14.7	01:57:43 ± 00:20:00	02:10:45 ± 00:19:44	8.4	4.9	11.8	9.9
2016	44 972	34.8	15.1	01:57:38 ± 00:20:00	02:11:16 ± 00:19:47	6.6	3.7	9.7	10.2
2017	42 252	34.5	13.9	01:57:27 ± 00:19:43	02:10:49 ± 00:20:03	6.0	3.8	14.3	12.6
2018	39 911	34.5	20.0	02:00:24 ± 00:21:17	02:14:40 ± 00:21:42	10.7	5.9	8.0	7.5
2019	33 134	34.0	19.4	01:59:58 ± 00:22:24	02:14:26 ± 00:22:05	11.1	6.7	8.2	5.4
Overall	**423 496**	**33.2**	**17.9**	**01:59:28 ± 00:20:14**	**02:12:33 ± 00:20:15**	**10.2**	**5.4**	**10.1**	**9.2**

[9] https://imbalanced-learn.org/stable/

The average runner in our analysis started the race at a faster pace than they could maintain, and gradually slowed down by each 5 km split, until the 20 km mark, when they managed to increase their speed with the goal in sight (Figure 1). Runners who made severe pacing errors displayed the same pacing pattern, but with an even faster start and larger drop in pace between 10-20 km. Runners who managed a negative split started slower in the first 5 km to then maintain a very even pace until the finishing sprint.

Most runners lost time in the second half of the race (Figure 2). An average runner, finishing in 120 minutes, lost just over 4 minutes on the second half. The fastest runners, finishing in less than 90 minutes lost less, on average around 1:30 minutes. In the groups with slower finishing times there were much more spread, but in general, slower finishers lost more on the second half. Note that among the very fastest, it seems few ran a negative split, possibly due to race tactics.

Average split differences were very similar between male and female runners. Grouping by finishing time however showed that female runners generally lost less time, see Appendix (Figure A1). Older runners (50+) had slower average finish times and larger split differences. Grouping by finish time showed no differences except among the slower runners (finish time > 150 minutes), where the younger age-groups in fact lost more time see Appendix (Figure A2).

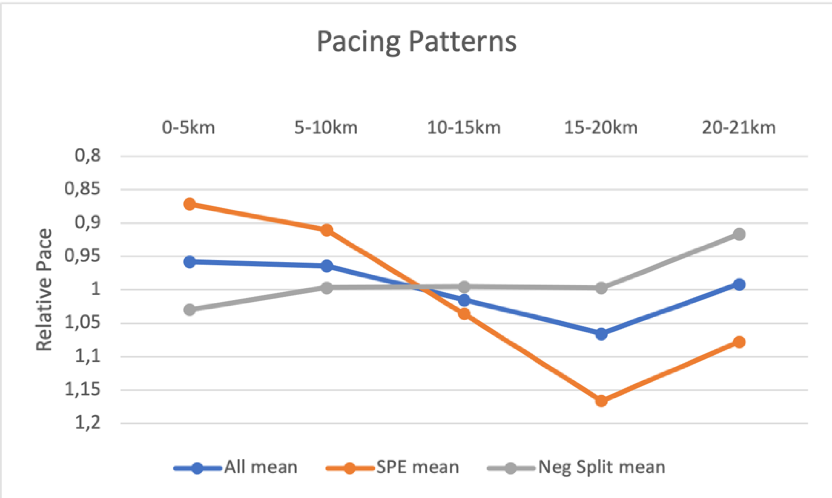


Figure 1: Relative pace for each segment.

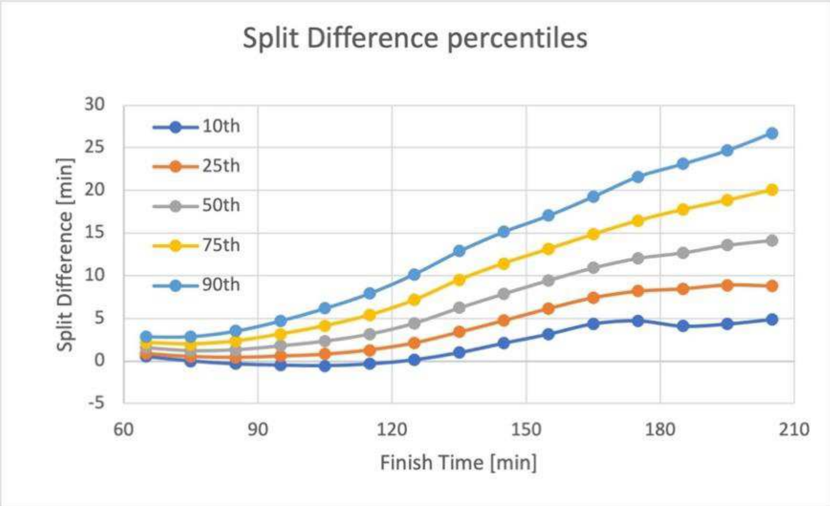


Figure 2: Average split difference (time lost on second half) as a function of finish time (all runners) showing 10th, 25th, 50th, 75th and 90th percentile in finish time grouped at 10-minute intervals.

Next, we investigate which runners, based on sex, age, and finishing time (as a proxy for runner fitness), applied a pacing strategy with a negative or equal split, and conversely, which runners experienced severe slowdowns.

Sex. Male runners were twice as likely to make a SPE: 10.2% of did so compared to just 5.4% of female runners (OR = 2.0; $p < 0.001$), see Table 1. Most runners slowed down during the second half of the race (Figure 1), but among runners managing a negative or equal split, male and female runners performed similarly: 10.1% of male runners and 9.2% of female runners (OR = 1.1; $p < 0.001$).

Age. Gothenburg Half Marathon is open for participants aged 17 and above, with most runners between 30-49 years of age, see Table 2. Age information was missing or incorrect for 3173 datapoints, which were excluded from analysis.

Table 2. Percentage of runners experiencing severe pacing errors or managing a negative split by age group.

Age	**#Runners**	**%Female**	**% SPE**		**% Neg Split**	
			M	**F**	**M**	**F**
17 – 29	89 032	44.1	13.1	7.0	14.7	12.3
30 – 39	125 484	32.9	10.2	4.7	11,6	10,1
40 – 49	124 275	31,3	9.0	4.2	9.0	7.7
50 – 59	62 261	26.5	9.3	5.5	6.3	4.7
60 +	19 272	17.3	9.9	5.4	4.4	3.0

The increased risk of SPEs for male compared to female runners was consistently high across all age groups: $1.74 \leq OR \leq 2.29$; $p < 0.001$ (Figure 3). For both sexes, the youngest runners (17-29 years old) were most likely to make SPEs, while the 40–49-year-olds were least likely (female: OR = 0.59; male: OR = 0.66; $p < 0.001$). Differences between consecutive age groups within sex are statistically significant except for the female runners in their 50's vs. 60's (p = 0.74).

Younger participants were more likely to run a negative split, with males slightly higher than females consistently across age groups: $1.17 \leq OR \leq 1.49$; $p < 0.001$. This decreased for each older age group, (Figure 3), pairwise between consecutive age groups of same sex, female: $0.58 \leq OR \leq 0.80$; male: $0.68 \leq OR \leq 0.76$; $p < 0.001$. The younger age groups are where we expect to find the elite or near-elite runners, who have the experience and fitness level to keep a consistent pacing for a full half-marathon, but perhaps also many inexperienced recreational runners who start too fast and later experience severe slowdowns.

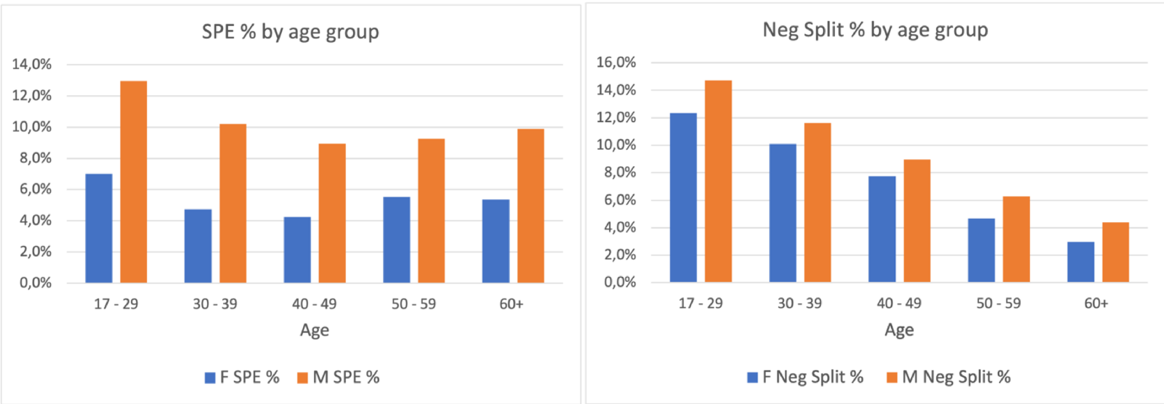


Figure 3: Proportion of male and female runners, by age group, who (left) made SPEs and (right) ran a negative split.

Finish Time. A larger proportion of runners made SPEs among those with slower than average finish times, both among males and females, see Appendix (Figure A3). For male runners, the proportion increased sharply for finish times above average (120 minutes), from less than 5% to over 30% among those finishing in over 150 minutes. The increase was less steep for female runners: for finish times below average (~135 minutes) less than 3% made SPEs, increasing to 25% for those who finished in over 180 minutes.

Conversely, the percentage of runners who managed a negative split was highest among those finishing in 75-104 minutes for males (15-16%) and 90-119 minutes for females (14%). This then dropped to 2-3% among the slowest runners. Note that among the very fastest group very few (<5%) ran a negative split, possibly because of race tactics and placement being more important than finish time.

Effect of Temperature. With higher temperature there was a trend towards both slower finish times and a larger proportion of SPEs. The average finish time and proportion of runners making SPEs was lower in the five coolest years studies (<18° C, small variation between years). In warmer years, many runners managed to compensate by reducing their tempo, see Figure 4 (female: $r^2 = 0.90$; male: $r^2 = 0.91$; $p < 0.001$). The difference in average finish time between the coldest (2012: 13,6° C) and the warmest (2013: 25° C) years was 7:36 minutes for female runners, and 9:18 minutes for males.

In warmer years, an increased proportion of male runners made SPEs: $r^2 = 0.85$; $p < 0.001$ (Figure 5). For female runners, temperature appeared to be less of a factor: $r^2 = 0.66$, $p = 0.004$. Regarding negative splits, the data fell into two clusters representing the five cooler years (<18°) and the five warmer years (> 18°). Runners were about twice as likely (female: $OR = 1.74$; male: $OR = 2.07$) to manage a negative split in the five cooler years, see Appendix (Figure A4).

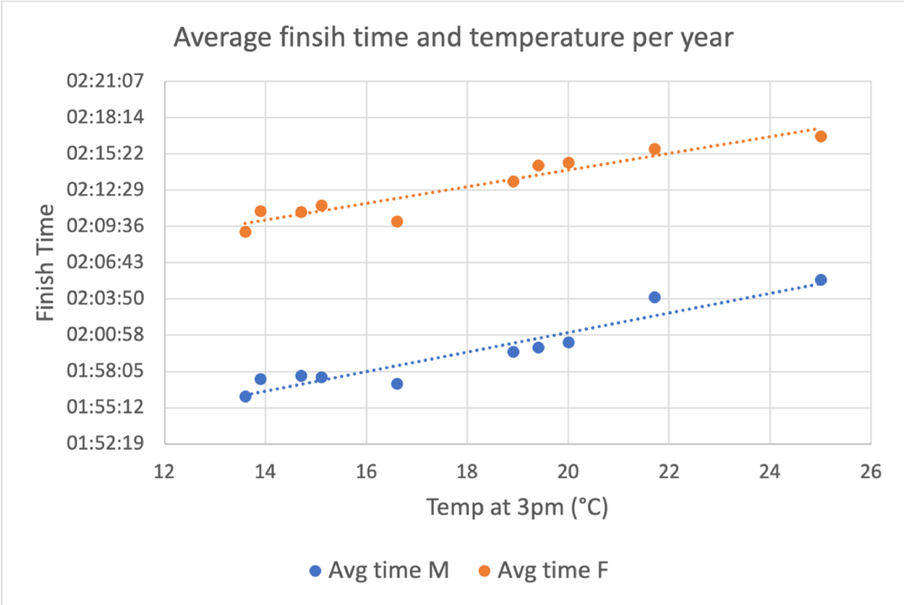


Figure 4: Average finish times for male and female runners. Each datapoint represent one year.

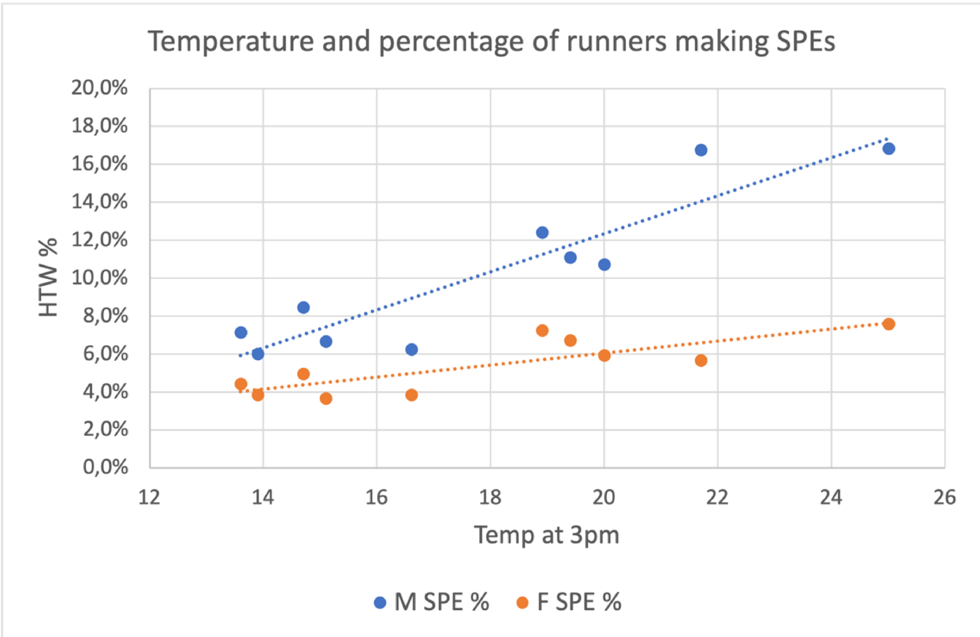


Figure 5: Temperature and percentage of male and female runners making SPEs per year. Each datapoint represent one year.

Predicting Finishing Time

For evaluation of the finishing time models, we compared the fixed pace baseline model with our linear regression and neural network models (Figure 6, left). We did a 5-fold cross validation with an 80-20% random split in training and test and computed the average mean absolute error (MAE) over all folds. We compared the predictions at each intermediate split, expecting the models with access to additional splits to perform better the closer to the finish we get. Both the linear regression models and the neural networks outperformed the baseline at all intermediate splits, with the neural network slightly outperforming the linear regression model at the 10 and 15 km splits. The largest absolute improvement over baseline was found at the 10 km split where the neural network had a 2:09 minute lower MAE than baseline.

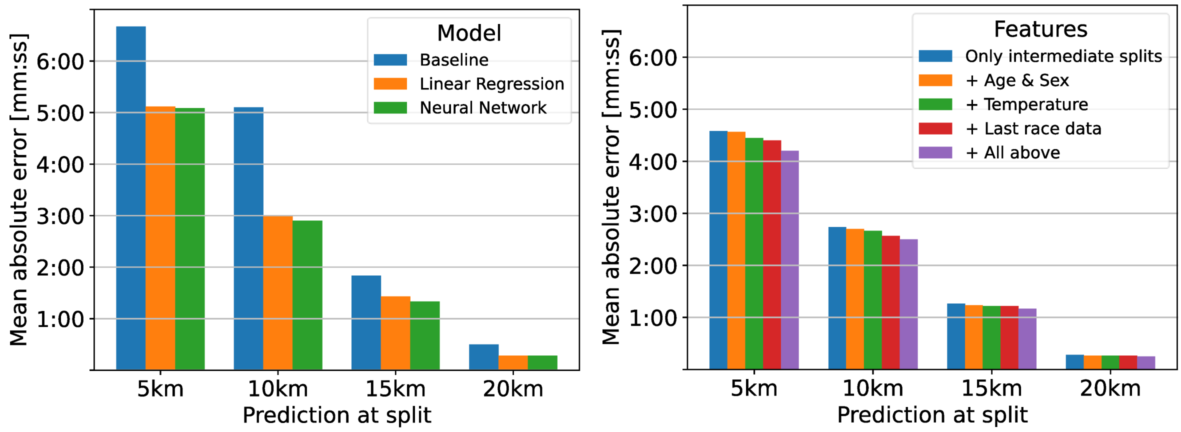


Figure 6. *Left:* Mean absolute error for the Baseline, Linear Regression and Neural Network models at all intermediate splits. *Right:* Mean absolute error for neural network models with and without additional features.

To further improve results, we added additional input features to the neural network based on factors from the exploratory data analysis: age, sex, and daytime temperature, as well as prior finishing time and pacing for repeat participants. To avoid data leakage, we used the 2019 race as our test set (25 443 samples), and data from earlier years for training (213 133 samples). Thus, the resulting prediction values cannot be directly compared to the method comparison analysis above. Adding additional features improved performance slightly, especially for predictions early in the race, where combining all features resulted in a MAE decrease of 33 seconds (Figure 6, right).

Predicting Pacing Categories

To learn about which runners are at risk of making an SPE, we need to also include some information approximating their capacity. However, as we did not have access to much data about the participants, beyond the results for the ten years under investigation, we opted to run these experiments on a subset of 238 576 data points, including only runners who had participated several times. The previous finish times were used as a proxy of the runners' presumed fitness and capacity. The input features selected through data analysis and initial experimentation were thus: age, sex, pace(0-5 km), pace(5-10 km), previous finish time and split ratio, number of previous races (a proxy for experience) and temperature.

As mentioned, the dataset is highly imbalanced, with 8.4% of data points belonging in the *SPE* and *Neg Split* categories respectively, and over 80% thus in the remaining category of runners making a moderate slowdown (labelled *Other*). As before, the data was divided into 80% for training (190 860 samples) and 20% for testing (47 716 samples), using stratified sampling to ensure a representative number of datapoints in each category, and downsampling the large *Other* class.

The balanced random forest model was fitted with a random parameter search (RandomizedSearchCV), based on average results for a 3-fold cross validation, maximizing the F1 score of the *Neg Split* and *SPE* classes. This was experimentally found to be optimal as it maximized the score for the difficult classes, with the results for the overall best model shown in Table 3. The model did a reasonable job of identifying the runners set to pace badly or well: at the 10 km split, it correctly identified 72% of those who would experience a severe slowdown, and 73% of those who managed a negative split (Table 3, *Recall* column). However, it did so at the cost of mislabeling many runners that in fact belong in the *Other* category, precision is only 0.20 and 0.22 respectively for the SPE and Neg Split categories, contributing to the relatively low F1 score.

Table 3. Metrics for classification of runners into those who made a severe pacing error (SPE), ran a negative split or did neither, predicted after the 10 km split.

	Precision	Recall	F1 Score
SPE	0.20	0.72	0.32
Neg Split	0.22	0.73	0.33
Other	0.92	0.47	0.62

From the confusion matrix (Table 4) we can see that the most common misclassifications were indeed runners belonging in the *Other* category. For the *SPE* and *Neg Split* categories, the most common mistake was to classify them as belonging to the *Other* category. We tried to explore if these misclassifications potentially lay close to the decision boundary, i.e., was close to make a SPE or manage a negative split but did not detect any clear patterns.

Table 4. Confusion matrix for the three pacing categories. Values along the diagonal show proportion of correct classifications for each class (recall).

	SPE	**0.72**	0.08	0.20
True Label	**Neg Split**	0.09	**0.73**	0.18
	Other	0.27	0.26	**0.47**
		SPE	**Neg Split**	**Other**

Predicted Label

Finally, by computing the permutation feature importance score, we measured which features were most important for the model. Unsurprisingly, the 5 and 10 km paces were by far the most important, followed by the previous finishing time. Sex, temperature, previous split ratio, and age was less important. The number of previous runs was not important for the model, and we also did not find any statistically significant patterns connecting this feature to an increased or decreased risk of SPEs. It could therefore have been removed from the machine learning model without affecting the results.

Discussion

Increased digitalization and availability of results and weather data allow for easier large-scale studies of pacing patterns of recreational races. Large datasets also allow training of machine learning methods. We used this methodology to answer whether machine learning could successfully: (RQ1a) predict the finishing time of runners at each split time and (RQ1b) identify, at the 10 km mark, which runners were at risk of making SPEs and which ones were following a low-risk pacing strategy. Via statistical analysis (RQ2), we investigated which features were potentially useful for such machine learning models, and if the trained models used these effectively.

For RQ1a, we found that a linear regression model performs better than the baseline model at every split time. An additional small improvement is obtained from the neural network model, but it seems the relationship between pace and finish time is largely linear, apart for a small subset of data which exhibits non-linear relationships. The neural network improvement was due to capturing these cases better. Adding the additional features from our statistical analysis to the neural network produced additional improvements in predictions, especially at early stages of the race, as expected.

Predicting which runners were at risk of making pacing errors (RQ1b), or conversely which runners were pacing within their limits proved to be a harder problem based on the current dataset. As no personal data (such as PR) was available, we reduced the dataset to only include those having participated at least twice, using their previous time as a proxy for ability. Fortunately, many runners participated multiple times, so this was not considered such a large limitation. Furtehrmore, the classification problem was unbalanced, with the *SPE* and *Neg Split* categories being much smaller than the category of runners losing a moderate amount of time (the *Other* category). We therefore used a balanced random forest model, which is designed for these cases. The model correctly identified 72-73% of the runner pacing badly and well respectively, but also mistook many that ought to have been in the *Other* category. As the training data undersampled the *Other* category, we believe the model learnt to expect a more even split between classes. Devising a better training procedure for this model was left as future work. We do however think that overclassifying SPEs is preferable over the converse, as the aim is to encourage recreational runners to keep a pace with low risk of overexertion. Regarding feature importance, the model learned to put more weight on the pace at the first two splits. This seem reasonable, as many runners who later make an SPE often have already started to slow at the 10 km mark, while runners pacing well are more even at this stage. Previous finish time was also a good indicator, consistent with the data analysis, as most SPEs were among slower runners. Contrary to results from the data analysis, the model did not put as much relative importance on sex and temperature. A possible explanation could be the low proportion of female runners in the dataset (only around 30%), and the fact that there were only a few years with exceedingly high temperatures. Balancing the dataset with respect to these features could have had a positive effect.

In addition to our machine learning results, our statistical analysis on this very large dataset over 10 years provided additional support for patterns seen in smaller studies. We found a higher proportion of male and younger runners making SPEs (age-group 17-29 years), and the lowest percentage among middle aged female runners (40-49 years), which is consistent with previous smaller studies on the marathon and half-marathon distance (Deaner et al., 2015; Smyth, 2021; Cúk et al., 2020, 2019). Carlström et al. (2019) found increased incidence of cases needing ambulances in years with temperatures above 17°C. The same patterns are mirrored among the runners that make SPEs during the race, with the same notable peaks in 2010 and 2013, see Appendix (Figure A5). This supports the conclusion of Carlström et al. (2019) that low-risk temperatures for half-marathons range between 13 – 18°C.

Limitations

Our definitions of good and bad pacing are estimates and do not include any personal metrics such as heart rate, hence the reasons for slowdowns are unknown. Similarly, the threshold for what constitutes a SPE could be tweaked. Still, we believe these definitions served as a good enough proxy for revealing trends in pacing, also seen in other studies. With access to more fine-grained data e.g., GPS traces (Berndsen et al., 2020), HR monitors and training history, the models can be made more exact. However, this requires additional data collection, and introduces the risk of skewing data towards ambitious runners, who are more likely to carry appropriate devices and record their training history.

We note that our dataset contains the same individuals running multiple times. We choose not to filter out repeat participants, as it would be difficult to decide which results to drop, and which to keep. Many runners may also have participated in years before those covered in our dataset.

Including repeat participants was also useful for the machine learning models, where prior results could be used as a proxy for fitness level.

Conclusion

The Gothenburg Half Marathon is one of the words largest and attracts many recreational runners. Through analysis of a large publicly available dataset of 10 years of results and split times, we showed that there is room for improvements in pacing as most runners slowed down throughout the race. Our goal with this work was to investigate if machine learning could help recreational runners with pacing, where we considered two tasks: predicting finishing time and identifying which runners were at risk of making a severe pacing error. We have taken a public health perspective, emphasizing running a safe race which avoids overexertion, rather than necessarily optimizing finish time under all conditions. We demonstrated improved accuracy over the current baseline on predicting runners finishing times, especially at the early stages of the race. Our model also demonstrated reasonable success in identifying which runners were at risk of experiencing a severe slowdown after the 10 km split, and conversely which runners were set to pacing evenly, with low risk of overexertion. As the finish time model was more successful than the pacing category prediction, treating the latter as a regression problem seems more suitable. Predicting how much time will be lost/gained as a numeric quantity, and comparing to a threshold was however left as future work.

Future work towards developing personalized pacing aids should investigate improving accuracy by including additional personal data, e.g., heart rate, training history and GPS. Until then, we expect our results may guide race organizers and recreational runners to mitigate common risks and assist in running a more enjoyable half-marathon.

References

Abadi, M. et al. (2016). TensorFlow: a system for large-scale machine learning. *In Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation (OSDI'16). USENIX Association, USA, 265–283.* Software available from tensorflow.org.

Abbiss, C., & Laursen, P. (2008). Describing and understanding pacing strategies during athletic competition. *Sports Medicine*, 38(3), 239-52.

Berndsen, J., Lawlor, A., & Smyth, B. (2020). Exploring the wall in marathon running. *Journal of Sports Analytics,* 6(3), 173 - 186.

Buitinck, L., Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, et al. (2013). API design for machine learning software: experiences from the scikit-learn project. *In Proceedings of the European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases,* 108–122

Buman, M. P., Omli, J. W., Giacobbi, P. R., & Brewer, B. W. (2008). Experiences and coping responses of "hitting the wall" for recreational marathon runners. *Journal of Applied Sport Psychology*, 20 (3), 282-300. Retrieved from https://doi.org/10.1080/10413200802078267

Carlström, E., Börjesson, M., Palm, G., Khorram-Manesh, A., Lindberg, F., Holmer, B., ... Thorsson, S. (2019). Medical emergencies during a half marathon race - the influence of weather. *Int. Journal of Sports Medicine*.

Deaner, R., Carter, R., Joyner, M., & Hunter, S. (2015). Men are more likely than women to slow in the marathon. *Med Sci Sports Exerc*., 47(3), 607-16.

Ely, M. R., Cheuvront, S. N., Roberts, W. O., & Montain, S. J. (2007). Impact of weather on marathon-running performance. *Medicine and science in sports and exercise*, 39(3), 487-493.

Khorram-Manesh, A., Löf, T., Börjesson, M., Nilson, F., Thorsson, S., Lindberg, F., & Carlström, E. (2020). Profiling collapsing half marathon runners — emerging risk factors: Results from Gothenburg half marathon. *Sports*, 8 (1). Retrieved from https://www.mdpi.com/2075-4663/8/1/2

Knechtle, B., & Nikolaidis, P. T. (2018). Sex- and age-related differences in half-marathon performance and competitiveness in the world's largest half-marathon - the GöteborgsVarvet. *Research in sports medicine*, 26(1), 75-85.

Lee, D., Brellenthin, A., Thompson, P., Sui, X., Lee, I., & Lavie, C. (2017). Running as a key lifestyle medicine for longevity. *Prog Cardiovasc Dis*.

Lemaître, G., Nogueira, F., Aridas, C.K. (2017) Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research* 18(17), 1–5.

Lüning, H., Mangelus, C., Carlström, E., Nilson, F., & Börjesson, M. (2019). Incidence and characteristics of severe exercise-associated collapse at the world's largest half-marathon. *PLOS ONE*, 14 (6), 1-9. Retrieved from https://doi.org/10.1371/journal.pone.0217465

March, D. S., Vanderburgh, P. M., Titlebaum, P. J., & Hoops, M. L. (2011). Age, sex, and finish time as determinants of pacing in the marathon. *Journal of Strength and Conditioning Research*, 25(2).

Nikolaidis, P. T., Cúk, I., & Knechtle, B. (2019). Pacing of women and men in half-marathon and marathon races. *Medicina*, 55 (1). Retrieved from https://www.mdpi.com/1648-9144/55/1/14

Roelands, B., de Koning, J., Foster, C., Hettinga, F., & Meeusen, R. (2013). Neurophysiological determinants of theoretical concepts and mechanisms involved in pacing. *Sports Medicine*, 43(5), 301–311. Retrieved from https://doi.org/10.1007/s40279-013-0030-4

Schwabe, K., Schwellnus, M. P., Derman, W., Swanevelder, S., & Jordaan, E. (2014a). Medical complications and deaths in 21 and 56 km road race runners: a 4-year prospective study in 65 865 runners - SAFER study I. *British Journal of Sports Medicine*, 48, 912 -918.

Schwabe, K., Schwellnus, M. P., Derman, W., Swanevelder, S., & Jordaan, E. (2014b). Older females are at higher risk for medical complications during 21 km road race running: a prospective study in 39 511 race starters - SAFER study III. *British Journal of Sports Medicine*, 48, 891 - 897.

Smyth, B. (2021). How recreational marathon runners hit the wall: A large-scale data analysis of late-race pacing collapse in the marathon. *PLoS ONE*, 16(5), 1-25. Retrieved from https://doi.org/10.1371/journal.pone.0251513

Trubee, N. W., Vanderburgh, P. M., Diestelkamp, W. S., & Jackson, K. J. (2014). Effects of heat stress and sex on pacing in marathon runners. *Journal of strength and conditioning research*, 28(6), 1673-1678.

Cúk, I., Nikolaidis, P., & Knechtle, B. (2020). Sex differences in pacing during half-marathon and marathon race. *Research in Sports Medicine*, 28(1).

Cúk, I., Nikolaidis, P., Markovic, S., & Knechtle, B. (2019). Age differences in pacing in endurance running: Comparison between marathon and half-marathon men and women. *Medicina (Kaunas)*, 55(8).