Abstract—Millions of people run. Movement scientists investigate the relationship of running kinematics to fatigue, injury, or running economy mainly using optical motion capture. It was found that running kinematics are highly individual and often cannot be summarized by single variables. We thus present a data-driven approach for movement analysis using wearable technology, combining statistical features and machine learning techniques, which allows to identify non-linear, complex relationships. Wearable technology enables running kinematic analysis to a broad mass in unconstrained environments. 20 runners wore 12 sensor units during two experiments: an all-out test and a fatiguing run. We used a Support Vector Machine (SVM) to distinguish skill level groups and achieved an accuracy of 76.92% with an acceleration sensor on the upper body. Sensor positions were ranked according to the movement change with fatigue using a feature selection. This ranking was consistent with visual annotations of a movement scientist. We propose a quantitative measure of movement change using a principal component analysis (PCA) and found an average correlation of 0.8369 for all runners with their perceived rating of fatigue.

Keywords—machine learning; measurement; wearable sensors

I. INTRODUCTION

The increasing participation in running in the 70s initiated an increase in running-related research [1] and did not break down to date. In contrast, in 2010, alone in the U.S. 507 000 runners finished a half-marathon, which set a new record. Approximately 30 million Americans run in total, including recreational and competitive runners [1]. Despite its popularity, running is considered to be a high injury risk sport with 66% of runners being injured in 2009 [2]. Reasons for injuries are manifold. Amongst others, fatigue was found to be a contributing factor to an increased injury risk [2], [3]. Additionally, running economy was found to be reduced with fatigue. However, the relationship of fatigue, muscle damage, and running economy to running kinematics is still not very well understood [4], [5].

State-of-the-art systems for measuring movement of runners are cameras or marker-based systems. From the captured movement, kinematic parameters (e.g. joint angles or temporal parameters) are extracted. These parameters are tested against hypotheses, e.g. the relationship to performance, mostly using statistical tests. However, running kinematics were found to be highly dependent on the individual and often it cannot be identified a single parameter linked to the hypotheses [6]–[8]. Despite their measurement accuracy, the main drawback of such systems is that they are constrained to an instrumented environment, since treadmill running is comparable to but not equivalent to overground running [9]. Additionally, the high cost of such systems and the high processing time mainly limits its use to elite athletes and research. With the emerging miniaturization and improvements in sensing accuracy, body-worn sensors are considered as a new modality for movement analysis in sports. They do not require a specific sensing environment and thus allow for unconstrained monitoring in the field.

In previous work we were able to extract established kinematic parameters from body-worn sensors for the assessment of skill level, technique, and fatigue [10], [11]. The results were consistent with movement scientist’s findings, confirming the use of body-worn sensors to substitute camera-based systems. While the analysis of the established kinematic parameters provides a powerful tool for movement analysis, these parameters often depend on the sensing modality, position, and performed movement [8]. Additionally, the analysis of single parameters does not identify high-dimensional non-linear relationships.

In this work, we thus explore the use of statistical features and machine learning approaches following a data-driven analysis for movement analysis in running. For the measurement of athletes’ movements, we used inertial measurement units (IMU) that were specifically optimized for unobtrusive movement monitoring in the field. For the data collection, we performed two experiments: runners of different skill level running at the same speed to investigate differences of running kinematics with skill level, and runners performing an exhausting run on an outside track to identify and track kinematic changes with fatigue. We thus investigate the potential of a data-driven approach for movement analysis in running and address the following research questions:

- Can statistical features calculated from wearable sensors be used to distinguish skill level groups?
- Can we provide a method to identify and track an individual runner’s movement changes with fatigue?
- How does the data-driven analysis compare to the state-of-the-art approach using kinematic parameters?

II. RELATED WORK

Of the numerous works on running we would like to focus on findings related to skill level and fatigue. To investigate
fatigue in running, the authors in [12] videotaped 8 subjects running on a treadmill before and after the completion of a marathon. When testing parameters measured pre-against post-marathon collected from the video recordings, they found that the kinematics changed with fatigue. However, the observed changes were highly individual. Authors in [3] found that an imbalance in the contraction of the shank muscles developed with fatigue, which leads to a higher risk of stress injuries. Step frequency was found to slightly increase during marathon [13]. Contradictory to that, Mizrahi et al. found that step frequency decreased over time [14]. [15] performed a study with 10 recreational runners performing a short high-intensity run to fatigue. Runners were found to perform higher knee flexion and rearfoot eversion (the ankle bended inwards) at heel impact. They concluded that the altered kinematics may have resulted in increased metabolic costs. Contrary to this finding, it was found that running kinematics did not change after a brief exhaustive run in a recent study [2]. In general, it was found that the observed changes often varied and were sometimes contradictory to previous studies [6]. This might be explained by differences in the experiment protocol such as duration or intensity of the run, but in most studies considerable inter-individual differences have been observed. Some runners might be potentially more sensitive to fatigue while others remain a rather constant movement pattern [6].

Considering running technique, Kenyan distance runners were found to perform very short foot contact duration, which may be related to their better running economy [16]. An analysis of kinematic and kinetic differences between runners of different performance levels identified several biomechanical variables, establishing the importance of biomechanical influences on running economy [7]. However, the authors concluded that it appears that no single variable or small subset of variables can explain differences in running economy between individuals but rather that economy is related to a weighted sum of the influences of many variables [7].

To allow for modeling of such non-linear, complex relationships between input and output data, the use of machine learning and pattern analysis techniques was introduced. Following this approach, authors in [8] used AdaBoost to classify data of runners collected from optical motion capture and ground reaction force measurements. They were able to discriminate shod from barefoot running, runners’ gender, and injured from non-injured runners. [17] used Support Vector Machines (SVMs) to distinguish young from elderly runners based on optical motion capture data.

The use of body-worn sensors to analyze running kinematics seems especially interesting because of the observed individuality of running kinematics. Additionally, it was found that treadmill running differs from overground running [9], [18], concluding that even for individual assessment of running kinematics on a treadmill to assess e.g. shoe or shoe orthotics may possibly lead to inadequate conclusions about overground running [18]. However, to date, there are still very few sensor-based approaches to kinematic analysis in running. Authors in [19] introduced an embedded classification of surface in a running shoe in cooperation with adidas. The shoe comprises a hall-sensor that was used to measure the deflection of the heel. They achieved a person-independent classification rate of over 80%. A drawback of the proposed system is that it only works for heel runners and not for toe or midfoot runners. [20] propose a pressure insole to monitor running technique. An ambitious recreational runner completed a prolonged run within a validation study. They showed that foot contact duration increased over time, possibly indicating fatigue.

## III. Methods

### A. Measurement Device

For the data collection we used ETHOS units [21]. ETHOS is an unobtrusive inertial measurement unit (IMU) specifically optimized for long-term recordings in unconstrained environments. Each unit consists of a 3D accelerometer, a 3D gyroscope, and a 3D magnetic field sensor. Connectivity is provided by an integrated ANT+ module and USB interface. The elongated design (WxLxH = 14 mm × 45 mm × 4 mm) is optimized for the attachment along human body limbs. We used flat and round housings (compare Fig. 1), the round housing unit weighs 27 g, the flat 22 g including the battery. Data were sampled at 100 Hz and stored to a local microSD card for subsequent offline analysis. Temporal alignment of simultaneously recorded data was guaranteed by a dedicated hub that synchronized the on-board real time clock (RTC) of attached ETHOS units.

![Image of sensor units](image)

**Figure 1.** Round (a) and flat (b) housing type sensor units. The housing each comprises an ETHOS unit, a battery, and a switch.

### B. Experimental Procedure

For the investigation of our hypotheses, we performed two experiments. For the discrimination of skill level groups, we recorded a dataset in which all runners performed a standardized all-out test. During this test, we controlled the runners’ speed with a treadmill, what ensured that observed differences in kinematics were not evoked by differences in speed. For the outside run an individual speed for each runner ensured that runners fatigued. The used experiment design was approved by an ethics commission.
1) All-out test: Runners performed a standardized all-out test on a treadmill to determine the individual runner’s fitness level. The maximal speed ($v_{\text{max}}$) achieved at the end of such a test has been shown to correlate well with running performance [22]. Runners started running at 5.8 km/h. Speed was increased every three minutes by 1.4 km/h until runners aborted the test due to exhaustion. The reached speed then was their maximum aerobic speed. We used a combination of this maximum speed and weekly training kilometers to group subjects in skill level groups, compare Table II. For the comparison of skill level groups, we chose the data collected during the speed step of 11.4 km/h since this was the last step that all runners managed to complete.

2) Exhausting Outdoor Run: The outdoor run was performed on an outside track commonly used by runners. Runners were instructed to maintain a speed of 75% to 80% of their maximum aerobic speed, determined in the all-out test, to ensure fatigue. The track offered a circular shape with a total length of 550 m, which allowed for permanent supervision of the subjects by an assistant for safety reasons. To assess runners’ perceived fatigue, we collected fatigue levels according to BORG’s perceived exertion and pain scales [23] every 3 min. These can be used to rate the perceived fatigue in a range from 6 to 20, compare Table I. They are commonly used in sports and the scale has been shown to correlate with the objective fatigue [24].

During the experiment, a movement scientist annotated the observed changes of movement for each runner throughout the run. These annotations allowed us to label the change of movement and were organized in categories according to the limb. We defined the categories feet, upper legs, lower legs, trunk, and arms. When changes were observed, the movement scientist annotated them in the according category with a timestamp, and briefly annotated the form of which they were. Exemplarily annotations were “change from toe to midfoot running” in the category feet, or “more forward leaning starting in minute 20” in the category trunk.

Table I

<table>
<thead>
<tr>
<th>rating</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-8</td>
<td>extremely light</td>
</tr>
<tr>
<td>9-10</td>
<td>very light</td>
</tr>
<tr>
<td>11</td>
<td>light</td>
</tr>
<tr>
<td>12-13</td>
<td>moderate</td>
</tr>
<tr>
<td>14-15</td>
<td>hard</td>
</tr>
<tr>
<td>16-17</td>
<td>very hard</td>
</tr>
<tr>
<td>18-19</td>
<td>extremely hard</td>
</tr>
<tr>
<td>20</td>
<td>maximum exertion</td>
</tr>
</tbody>
</table>

BORG’S PERCEIVED EXERTION SCALE THAT WAS USED TO ASSESS PERCEIVED EXERTION DURING THE EXPERIMENTS: RATING AND ACCORDING DESCRIPTION.

C. Dataset

In total 23 healthy subjects (14 male, 9 female) participated in our experiments. They were aged between 23 and 50 years, with a mean of 33.8 years. During all experiments, runners were equipped with 12 ETHOS units on a runner that were worn throughout the experiments.

Table II

<table>
<thead>
<tr>
<th>skill level group</th>
<th>$v_{\text{max}}$ [km/h]</th>
<th>training [km/week]</th>
<th>subjects [#]</th>
</tr>
</thead>
<tbody>
<tr>
<td>beginner</td>
<td>&lt; 14</td>
<td>0 – 4</td>
<td>7</td>
</tr>
<tr>
<td>intermediate</td>
<td>14 – 16</td>
<td>20 – 40</td>
<td>7</td>
</tr>
<tr>
<td>advanced</td>
<td>&gt; 16</td>
<td>30 – 150</td>
<td>6</td>
</tr>
</tbody>
</table>

SUBJECTS, THEIR MAXIMUM VELOCITY IN THE ENDURANCE TEST ($v_{\text{max}}$), THEIR AMOUNT OF TRAINING, AND THEIR ASSIGNMENT TO A PERFORMANCE GROUP

Figure 2. Front (a) and back (b) view of the positioning of the 12 ETHOS units on a runner that were worn throughout the experiments.

IV. DISCRIMINATION OF SKILL LEVEL GROUPS

In previous work we found that established kinematic parameters calculated from sensor data gathered at the foot and the hip can be used to distinguish experienced from unexperienced runners [10]. To investigate if statistical features calculated from wearable sensor data can be used to discriminate between the three skill level groups, we used data collected during the all-out test. Runners thus all ran at the same speed, ensuring that the identified differences are not attributed to speed.
A. Feature Extraction

Following the common approach in data-driven classification, we perform feature extraction, feature selection, and classification. Features were calculated on a 5-s sliding window with 2 s overlap, to include several step cycles within one window. The following features were calculated for acceleration and rate of turn of each sensor node: mean value, standard deviation, skewness, kurtosis, range (max - min value), first frequency component, number of values below mean. Opposed to optical systems, sensor displacement is an issue when using body-worn sensors in unconstrained environments. Even though sensors were securely fixed to runners’ bodies, we calculated the features from the magnitude of the sensor modalities instead of the individual axes, to account for possible sensor displacement. From the feature extraction, we obtained a 98-dimensional feature vector.

B. Feature Selection and Classification

To classify runners into skill level, we use the grouping according to Table II as ground truth. For the discrimination, we combine feature selection and classification with a leave-one-subject-out scheme. To avoid overfitting, the common approach is to leave one subject out for testing and part the remaining data in one set for feature selection and one set for training. The classifier was then trained using the training subset and the features obtained from the feature selection subset and tested on the remaining data. To calculate the classification accuracy, we averaged over the accuracies obtained from the individual classifications.

For the feature selection, we first discarded features whose correlation was greater than 0.8 with any other feature. Afterwards, we chose a filter method called minimum redundancy maximum relevance (mRMR) [25] over the commonly used wrapper approaches because of the lower computational costs. The goal of this algorithm is to rank the features according to the representation of the dataset based on the ground truth labels while features are chosen to be mutually as dissimilar to each other as possible. A combination of the first $k$ features is then considered as the selected features. However, the algorithm does not provide the number of needed features $k$ to maximize the classification performance. We thus chose $k$ after a parameter sweep. For classification, we chose a support vector machine (SVM).

Following this approach, we achieved a mean classification accuracy of 75.61% using 10 features and data from all sensors. With a majority voting to assign a single label to each runner, the accuracy increased to 76.92%. This implies that 3 runners of the 13 were misclassified. Since the features were chosen for each leave-one-subject-out subset they varied across subjects. However, 4 features were selected within all selected feature sets of 10 features, listed in Table III.

Table III suggests to use solely the sensor on the upper back for the classification task. This finding was confirmed when reducing the number of sensors. The highest accuracy of 69.04% (76.92% with majority voting) was achieved

<table>
<thead>
<tr>
<th>sensor position</th>
<th>feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>upper back</td>
<td>mean acceleration</td>
</tr>
<tr>
<td>upper back</td>
<td>standard deviation acceleration</td>
</tr>
<tr>
<td>upper back</td>
<td>number of samples below mean acceleration</td>
</tr>
<tr>
<td>foot</td>
<td>number of samples below mean signal value</td>
</tr>
</tbody>
</table>

when solely using the sensor from the upper back with 3 features. Figure 3 depicts a scatter plot of these.

Figure 3. Scatter plot of features collected from the upper body. With a SVM we achieve a classification accuracy of 76.92% for the discrimination of skill level with this single sensor.

Using the majority voting thus allows to discriminate the three skill level groups using a single acceleration sensor on the upper back.

V. MOVEMENT CHANGES WITH FATIGUE

Despite the increasing injury risk with fatigue [2], [3], research in movement sciences found that the changes in running kinematics with fatigue are highly individual [6]. It thus seems impossible to define a single sensor position to measure fatigue across all runners. Usually observations by trainers are used to support athletes. They identify their weaknesses and support by giving e.g. instructions on strength training. Since not all runners have regular support by a trainer, we aim at identifying movement patterns that change with fatigue automatically for individual runners. In a first step we thus introduce automatic feature selection for an importance ranking of the different sensor positions, compare Section V-A. We then provide a method to quantify movement change with fatigue (Section V-B).

A. Identification of movement changes with fatigue

As explained in Section III-B, movement changes during the fatiguing run were annotated by a movement scientist. To detect these changes automatically, we used statistical features calculated from the sensor data (compare Section IV-A) and the feature selection scheme described in Section IV-B using mRMR. To identify features that changed over the course of the run, we fed data from the first and the last 5 min in the mRMR algorithm and thus got a ranking of features for the discrimination of the start from the end of the run. To identify areas of movement change we investigated the 50 best features of this ranking. For each sensor position we defined an importance count of the according sensor position. The annotations of the movement scientist during the run (see Sec. III-B2) provided labels if changes were observed for
specific categories (body parts) or not. We colored bars red when changes were observed for the individual categories, as depicted in Figure 4.

The figure illustrates that the importance count was consistent with the annotations of the movement scientist. However, for the most experienced runners (3 most right runners in the last row of Figure 4) some sensor positions achieved a high importance count even though no changes were observed by the movement scientist. We will briefly discuss the results using three examples:

**Beginner (***):** The importance count of a beginner runner is indicated with (**) in Fig. 4. The movement scientist annotated increasing arm movement for this runner. The finding from the data-driven approach was consistent with this ground truth annotation, as it can be seen by the high importance count of the arm sensors.

**Advanced runner (***):** Runner (**) was an advanced runner, regularly competing in triathlons. The movement scientist’s annotation of increasing forward leaning of the trunk agrees with the finding from our automatic method.

**Expert runner (**):** Runner (***) was an expert runner who completed the 45 min with a speed of 16.4 km/h. From the visual annotation no prominent changes during the run occurred, only a very slight increase in trunk forward leaning. This is consistent with our feature selection since no importance count is very prominent. However, the slight change in upper body posture is also visible in the importance count.

### B. Quantification of movement changes with fatigue

During the experiment, the perceived fatigue level was gathered with the established BORG scale, as described in Section III-B. However, since this measure is highly subjective, we were interested in obtaining an objective quantitative measure from the recorded movement.

To reduce the feature dimension, we used a principal component analysis (PCA). Figure 5 depicts exemplarily the first two principal components of an intermediate runner. The data points are colored according to the respective BORG value. The plot indicates that the scatter cloud moves with increasing fatigue in PC space. To capture this movement, we calculated the distance between the start (mean of the first 2 min of the run) and the individual point in this PC space using the first 3 principal components.

This distance is depicted together with the BORG values in Figure 6. The plot suggests that this measure is correlated with the perceived fatigue level. We thus calculated the correlation coefficient of the filtered distance and the BORG values and achieved a mean value of $\bar{r} = 0.8369$, $p < 0.05$ across all subjects.

Choosing for each subject the sensor that achieved the highest importance count according to Figure 4 yields an average correlation value of $\bar{r} = 0.8688$. Since the kinematic changes with fatigue are highly depending on the individual, which is consistent with movement scientists’ findings [6], the correlation decreased when using data from a single sensor position across all subjects. The maximum was achieved for the upper body sensor with $r = 0.6407$.

### VI. Conclusion and Outlook

In this work, we presented the use of statistical features extracted from body-worn sensors for the discrimination of skill level groups. With a leave-one-subject-out scheme we achieved a classification accuracy of 76.92% using a single acceleration sensor on the upper back. To investigate individual kinematic changes with fatigue, we performed an experiment with 20 runners completing an exhausting run on an outside track. A ranking of the sensor positions according to the number of selected features identified most prominent movement changes, which were consistent with the visual annotations of a movement scientist. To quantify the change of movement, we performed a PCA and calculated a distance measure in the space of the first three principal components. This measure was significantly correlated with the perceived fatigue level ($\bar{r} = 0.8369$), possibly providing a quantitative measure of kinematic changes due to fatigue.

We concluded that a data-driven analysis using statistical features collected from body-worn sensor data shows high potential for movement analysis due to its ability to reveal complex, non-linear relationships of movement and skill level or fatigue. The use of body-worn sensors is especially important since research in movement sciences found that kinematic changes are highly depending on the individual.
Figure 4. A mRMR feature selection identified the features that changed most throughout the run. The histograms (one per runner) show the number of selected features per sensor position within the top 50 features. The findings from the histograms were consistent with annotations based on visual observations of a movement scientist. Bars are colored red (grey) when the movement scientist observed a change in this limb. (*) was a beginner runner who increased arm movement. (**) was an advanced runner who leaned more forward. (**) was an expert runner who did not change his movement visibly but slightly changed the upper body form.

With a feedback runners could work on improving their weak areas, e.g. training the lower back muscles to decrease the risk of injury. However, such a feedback remains to future work. Additionally, further investigations on repeatability and experiments including more subjects will be used to further validate the proposed methods.

REFERENCES