Article

Toward the Personalization of Biceps Fatigue Detection Model for Gym Activity: An Approach to Utilize Wearables' Data from the Crowd

Mohamed Elshafei [ID], Diego Elias Costa [ID] and Emad Shihab *

Department of Computer Science and Software Engineering, Concordia University,
Montreal, QC H3G 1M8, Canada; m_lshafe@encs.concordia.ca (M.E.); diego.costa@concordia.ca (D.E.C.)
* Correspondence: eshihab@cse.concordia.ca

Abstract: Nowadays, wearables-based Human Activity Recognition (HAR) systems represent a modern, robust, and lightweight solution to monitor athlete performance. However, user data variability is a problem that may hinder the performance of HAR systems, especially the cross-subject HAR models. Such a problem may have a lesser effect on the subject-specific model because it is a tailored model that serves a specific user; hence, data variability is usually low, and performance is often high. However, such a performance comes with a high cost in data collection and processing per user. Therefore, in this work, we present a personalized model that achieves higher performance than the cross-subject model while maintaining a lower data cost than the subject-specific model. Our personalization approach sources data from the crowd based on similarity scores computed between the test subject and the individuals in the crowd. Our dataset consists of 3750 concentration curl repetitions from 25 volunteers with ages and BMI ranging between 20–46 and 24–46, respectively. We compute 11 hand-crafted features and train 2 personalized AdaBoost models, Decision Tree (AdaBoost-DT) and Artificial Neural Networks (AdaBoost-ANN), using data from whom the test subject shares similar physical and single traits. Our findings show that the AdaBoost-DT model outperforms the cross-subject-DT model by 5.89%, while the AdaBoost-ANN model outperforms the cross-subject-ANN model by 3.38%. On the other hand, at 50.0% less of the test subject's data consumption, our AdaBoost-DT model outperforms the subject-specific-DT model by 16%, while the AdaBoost-ANN model outperforms the subject-specific-ANN model by 10.33%. Yet, the subject-specific models achieve the best performances at 100% of the test subjects' data consumption.

Keywords: human activity recognition; wearable sensors; wearable sensor data; machine learning

1. Introduction

Fatigue is a natural phenomenon that describes physiological impairments or lack of energy caused by prolonged activities [1]. Lately, a plethora of works in the literature about athlete training and monitoring approaches aim to track body performance and measure muscle fatigue for two reasons: (1) fatigue is an important stamina indicator as previous studies show that it often occurs prior to muscle injuries in over-training [2–4] and (2) fatigue-induced injuries are one of the most active research topics in sports science due to their impact on athletes, such as substantial loss in muscle strength and flexibility [5].

Often reported as one of the earliest, the invasive approach is used to detect fatigue by measuring the lactic acid in the bloodstream to determine the maximal muscle effort that a person can maintain without risk of injuries [6,7]. Although such an approach requires puncturing the skin, it often provides accurate information about fatigue conditions and acetate levels in muscles. A previous study requires several blood samples from athletes for a blood lactate test to measure their muscle endurance that can provide accurate results up to 97% [8]. Another previous study that requires blood samples from marathon runners for creatine kinase (CK) tests shows that creatine kinase levels are significantly different

(p-value < 0.05) in muscles during running, which indicates the risk of skeletal muscle injuries [9]. Less painful but respiratory related, the cardio-respiratory approach monitors a person's fatigue levels by measuring their circulatory and respiratory systems' ability to supply oxygen to skeletal muscles during sustained physical exercise without risk of injuries [10,11]. Such an approach may require up to five pieces of equipment such as blood pressure cuff, Electrocardiograph (ECG), bicycle, mouthpiece, and saturation monitor. This setup is complex and requires technical expertise to run; hence, it is considered a drawback for daily living applications. However, it accurately estimates the maximum volume of oxygen consumption VO_2 during incremental exercises to detect fatigue. A previous study on the oxygen consumption in skeletal muscles measures the respiratory system's efficiency during sustained physical activity shows that oxygen consumption can provide an accurate indication ($>94\%$) of fatigue intensity [12]. Although the previous approaches may detect signs of fatigue precisely, they often require medical equipment and field training, which makes them too complicated for ordinary users. Recently viewed as the latest, the wearable approach is used to detect fatigue by utilizing sensory data and machine learning algorithms to estimate rates of perceived exertion (RPE) during physical exercises. After the rapid development of low-power wearables, such an approach often represents a modern way to monitor athlete performance and to prevent fatigue-induced injuries during training [13,14]. Previous work on marathon runners that uses data from the inertial measurement unit (IMU) to classify runners as being not-fatigued or fatigued shows that ROC Curves for the General Runners Models, trained using only the statistical features, range between (0.67 and 0.71) [15].

HAR applications in motion tracking and athletic training have become more prominent as wearable technology and machine learning techniques advance [16,17]. Yet, a common obstacle in such applications is having sufficient data to train the HAR models reliably [18,19]. Training HAR models using insufficient data limits their performance and may even make them impractical for their user base. One way to work around this obstacle is to collect data from a large pool of users and train a cross-subject HAR model. However, this does not guarantee an accurate performance from the cross-subject HAR model because even with sufficient data from a large pool of users, individuals may perform the same activity differently. This increases the inter-subject data variability, which hinders the performance of HAR applications [19]. The inter-subject data variability is often high in places where there is a diverse crowd of users with different physical traits [19,20]. A way to reduce the inter-subject data variability is to address each user separately by collecting data from the user of the HAR application to train a subject-specific HAR model. However, the cost of training a subject-specific HAR model is often prohibitive and requires labeled data from the user [21–23]. Therefore, there is an inherent trade-off between cross-subject models (cheaper but less accurate) versus subject-specific models (more expensive and more accurate). Furthermore, such a trade-off often exacerbates in specialized cases in HAR (e.g., muscle fatigue detection), where manual or semi-supervised labeling is usually required [24,25]. As a result, this increases the data cost in the case of the subject-specific models; or, if we want to spare that data cost, we will choose the less accurate option, the cross-subject models. In this work, we attempt to bridge the trade-off between cross-subject and subject-specific models. We propose a personalization approach that improves the performance of the cross-subject model while utilizing a small sample of the user's data.

To further explain our approach, recent studies show that the personalization of HAR cross-subject models can improve the models' performance while consuming a part of the test subject's data less than the subject-specific models [26,27]. The rationale behind the personalization is to prioritize the crowd's data from whom the test subject is similar. Therefore, we study the physical traits of the crowd to improve personalization of the cross-subject model by adding weights to the training data entries based on the inter-subject similarities between the test subject and the subjects from the crowd [28–32]. We believe that combining similarities measure and personalization can form a personalized model which has a better performance than the cross-subject model and consumes the

test subject's data less than the subject-specific models. Moreover, we further enhance the models' performance by combining hand-crafted features with some of the suggested ones from the previous studies [15,33,34]. The hand-crafted features are often specific toward the activities of interest, which the models have to detect. This further improves the weighting of the training data based on the similarities between training subjects, the crowd, and the test subject.

To evaluate our approach, we perform a case study on detecting biceps muscle fatigue during gym exercise. The study investigates the increasing inter-subject data variability in data patterns for the same exercise, e.g., bicep curls, among gym-goers due to their physical properties and movements. Sometimes, introducing a new member to the gym may increase the inter-subject data variability between the gym-goers if such a member performs exercises differently from the crowd, wherein this often hinders the HAR models' performance [35]. A previous study shows that using a subject-specific model to address each subject may overcome the inter-subject variability challenge, therefore improving the model's performance [36]. However, these models are known for their high demand for the subject's data which is its major drawback [18]. Moreover, these models often suffer from dramatic performance drops in real-time applications if the subject's data are unavailable or insufficient [37,38].

Detecting bicep muscle fatigue is important because bicep muscle is one of the most active skeletal muscles in the elbow joint [39,40]. Moreover, bicep fatigue injuries, such as muscle strain, may require up to 22 weeks of treatment [41,42], while tendon rupture results in a substantial permanent decrease in flexion and extension strength [43]. This may delay athletes' training schedules or their immediate withdrawal from competitions [44]. Therefore, given the detrimental effects of bicep muscle fatigue and the low performance of cross-subject detection models, we believe that further methods are needed to improve the performance of the detection models. In our previous work, fatigue detection models achieved accuracies of 98% and 88% for subject-specific and cross-subject models, respectively. However, we observed that as the inter-subject data variability increases, the gap between the performance of cross-subject and subject-specific models widens from 10% to 15% [45]. Therefore, in this work, we propose the personalization of cross-subject models as a solution that provides better performance than cross-subject models while requiring less data than subject-specific models. Moreover, we provide a set of eleven hand-crafted features that enhance the performance of personalized fatigue detection models.

In this work, we employ the personalization of the HAR models to improve the performance of the bicep fatigue detection models. In Figure 1, we show an overview of the approach used in this work. First, we select biceps concentration curls as the activity of interest for this work. Second, we start collecting our data from 25 participants by placing a 50 Hz Neblina inertial measurement unit (IMU) on their wrist. Meanwhile, on the other wrist, we place an Apple Watch Series 4 to measure their heart rate during the exercise. We choose their wrist because it is a primary measuring point for upper limb-related activities. Third, we manually label each repetition in our dataset as containing fatigue or not containing fatigue; then, we compute eleven hand-crafted features commonly reported in the HAR and fatigue-related works. Fourth, we measure two types of similarities, physical and signal similarities between the test subject and the crowd, and we use these similarities to add more weight to data collected from individuals in the crowd who are similar to the test subject. Fifth, we use the weighted dataset to train two machine learning models: Decision Trees (DT) and Artificial Neural Networks (ANN). Finally, we test these models using the test subject's data in detecting fatigue in biceps concentration curls.
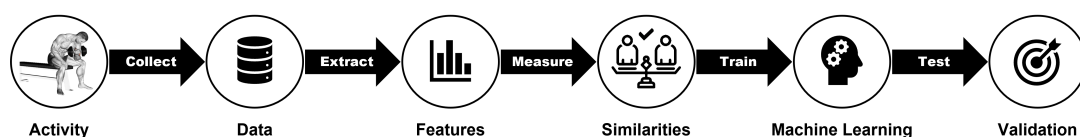

Figure 1. Overview of the similarity-based personalization approach for HAR.

Our first finding shows that it is preferable for the personalized fatigue detection models to depend on both physical and signal similarities with a tendency toward signal similarity. This finding is our first contribution where the eleven hand-crafted features aid the personalization approach in identifying individuals with similar biceps movement patterns to contribute more in the training dataset and hence improving the performance of fatigue detection models. Our second finding shows that the personalization approach improves the accuracy of the DT model by 5.89% and 3.38% for the ANN model. This finding is our second contribution where such improvements occur in personalized cross-subject models after prioritizing training data from the crowd based on the total similarity score, where individuals with high scores contribute more to the models' training dataset. Our third finding shows that we can further improve the accuracy of personalized models by adding a portion of the test subject's data to the models' training datasets. In this case, the personalized DT model achieves 16.0% higher accuracy than the cross-subject DT model while consuming 50.0% less of the test subject's data than the subject-specific DT model. Moreover, the personalized ANN model achieves 10.33% higher accuracy than the cross-subject ANN model while consuming 33.33% less of the test subject's data than the subject-specific ANN model. This finding is our third contribution where the personalization approach benefits from the additional data in the training set; yet, it maintains a comparatively lower test subject data consumption by 33.3% up to 50.0% compared to the subject-specific models.

The rest of the paper is structured as follows. Section 2 describes how we collect our dataset and label its entries, as well as illustrates our approach to feature extraction and measure similarities between the test subject and others from the crowd. In Section 3, we present our three experiments and research questions; then, we show the results. In Section 4, we discuss our findings. Section 5 concludes our paper.

2. Materials and Methods

This section describes the details of our dataset and its characteristics to represent a sample from the crowd. In addition, we demonstrate the processing steps applied to our dataset to provide a high-quality dataset for our study. Moreover, we extract the eleven features to detect biceps muscle fatigue. Finally, we describe our approach to measure the similarity between a test subject and the crowd.

2.1. Dataset Description

We ask twenty-five volunteers to perform concentration curls while we use the following tools to construct our dataset: (1) We attach a 50 Hz Neblina inertial measurement unit (IMU) to the volunteer's wrist to measure its acceleration and calculate the linear and angular velocities. Previous studies show velocity loss as an early indicator of muscle fatigue during resistance training, especially when blood lactate and ammonia accumulate in muscle tissues [46–48]. (2) We attach an Apple Watch Series 4 to the volunteer's opposite wrist to measure their heart rate during the exercise. (3) We provide the volunteers with a 4.5 kg weight dumbbell to perform concentration curls. (4) We provide the volunteers with Borg's scale presented in Table 1 to express their fatigue levels. Such a scale is often used as a subjective method to estimate the rate of perceived exertion (RPE), which expresses the fatigue intensity during an exercise. During data collection sessions, we ask each volunteer to complete 5 warm-up repetitions followed by 15 repetitions per set for a total of 5 sets per hand, as shown in Figure 2. Moreover, the volunteers report their RPE after each set, including the warm-up, yielding 6 RPE values per hand.

Table 1. Borg G.A. psychophysical bases of perceived exertion [49].

Perceived Exertion	Borg Rating	Examples
None	6	Reading a book, watching television
Very, very light	7 to 8	Tying shoes
Very light	9 to 10	Chores such as folding clothes that take little effort
Fairly light	11 to 12	Walking through a store (without speeding breath)
Somewhat hard	13 to 14	Brisk walking (mild effort and speeding breath)
Hard	15 to 16	Bicycling, swimming (effort and heart pounding)
Very hard	17 to 18	Intense activity but can be sustained
Very, very hard	19 to 20	Very intense activity that cannot be sustained
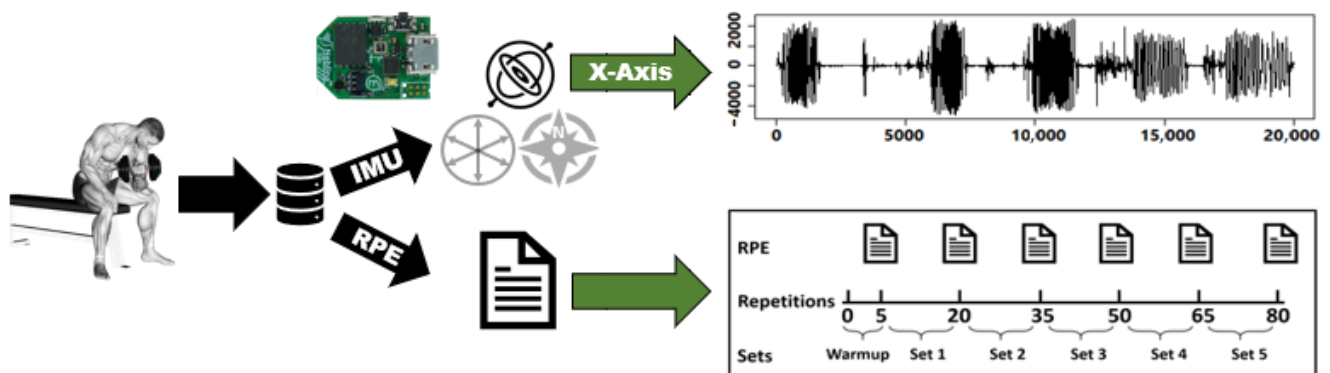
Figure 2. Visualization of data acquisition sessions of biceps concentration curl exercise. Rating of perceived exertion (RPE).

It is essential to explain the rationale behind the tools used in the data collection sessions, such as the Apple Watch, the 4.5 kg dumbbell, and the Borg's scale. Borg's scale ranges from 6 to 20, where by multiplying these values by ten, we can estimate the volunteer's heart rate during the exercise. For example, if a volunteer reports 13 on Borg's scale, we should expect to measure their heart rate around 130 to 140 by the Apple Watch. This serves as a way to strengthen the validity of the reported RPE by each volunteer. However, in the rare cases of dissimilarity between the Borg scale and the measured heart rates, we average the reported RPE with the measured heart rate converted to RPE, as similarly performed in previous work [50]. We use the 4.5 kg weight dumbbell in our work because of three reasons. The first reason is that several previous works have used medium-weight dumbbells ranging between 3.5 kg and 5.5 kg to study muscular strength and fatigue [51–53]. The second reason is that medium-weight dumbbells are often reported as the most commonly used dumbbells across gym-goers [54]. Third, in previous work, we found that the 4.5 kg weight dumbbell provides the best trade-off between number data points recorded in data sessions and time to reach fatigue during an exercise [55]. We use Borg's scale in our work because we believe that RPE is an appropriate marker of fatigue as previous studies within sport science have proven that RPE is capable of modeling a person's performance better in the real world compared to only heart rate monitoring [7,49,56].

To aptly evaluate our personalization approach, our dataset should contain sufficient data from a diverse set of users performing biceps concentration curls for four reasons. First, with enough data points, we can better capture bicep muscle fatigue, which helps us spot the variations of fatigue patterns among the volunteers. Second, a diverse dataset strengthens our work and better represents a sample from the crowd. The twenty-five volunteers are diverse in age, ranging between 20 and 46, as previous studies show that the selected age covers three distinct stages of athletes' performance: early, middle, and late, where athletes usually notice physical declines [57,58]. Moreover, the twenty-five volunteers are diverse in weight and height, ranging between 69–127 and 165–190, respectively. It

is important to have such variety, as physical characteristics such as weight and height affect arm movements and the severity of injuries, including fatigue-induced ones [59]. We calculated the Body Mass Index (BMI) for the twenty-five volunteers using the following formula [60]:

$$\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height(m)}^2}$$

The twenty-five volunteers are also diverse in BMI, ranging between 24 and 46, to include normal weight ($18.5 \leq \text{BMI} \leq 24.9$), overweight ($25 \leq \text{BMI} \leq 29.9$), and obesity ($30 < \text{BMI}$) [61]. The two black dots in Figure 3d are for two volunteers with outlier BMI values of 41 kg/m^2 and 46 kg/m^2 who are considered extremely obese [62]. Previous studies show that the relationship of BMI to injury risk is bimodal, where trainees with the lowest BMIs exhibit the highest injury risks for both genders and across all fitness levels [63,64]. In addition, the volunteers have no chronic diseases, no muscle or bone surgeries, and have been gym-goers for at least 1 year. Moreover, the volunteers are not on prescribed drugs or substances expected to affect their physical performance.
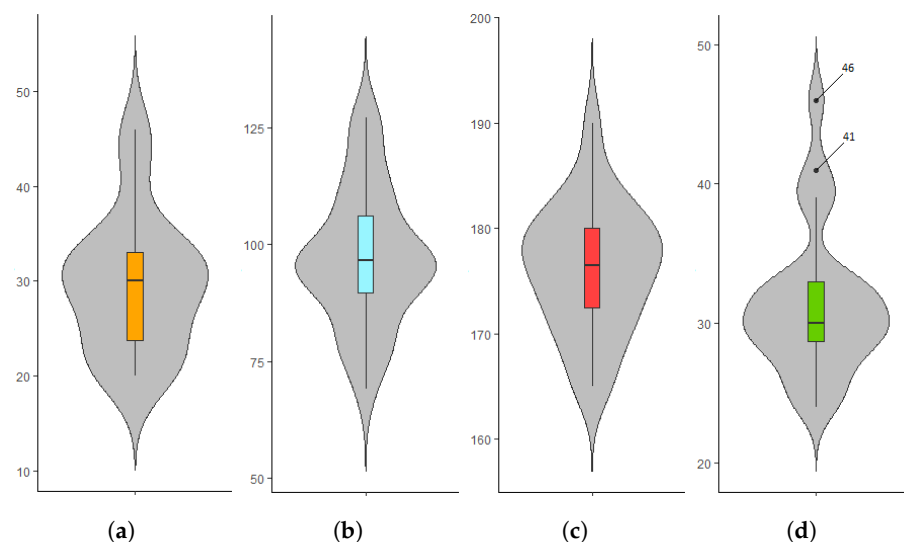

Figure 3. Boxplots to display the distribution of volunteers' age, weight, height, and BMI in our dataset. (**a**) Age (years). (**b**) Weight (kg). (**c**) Height (cm). (**d**) BMI (kg/m^2).

2.2. Dataset Processing

Initially, our dataset contains six signals from the two 3D-accelerometer's signals (x, y, z) and the gyroscope. We start by extracting and labeling the repetitions into fatigue and non-fatigue repetitions. Figure 4 shows an example of the fifth set of repetitions in its raw data form. In this example, the raw data are extracted from the gyroscope's x-axis along with their corresponding RPE values reported by the volunteer. A previous study on quantifying muscle fatigue suggests an RPE value of 16 as the threshold of true fatigue to estimate the declines in muscle strength during tasks [65]. Therefore, we extract and label each repetition manually according to the RPE values reported for the set, where we label repetitions with reported RPE values larger than 16 as fatigue and others as non-fatigue repetitions. The figure shows two distinct groups of non-fatigue and fatigue repetitions extracted from the set. The troughs indicate full extension, while peaks indicate full flexion. We repeat the same process for the remaining 4 sets.
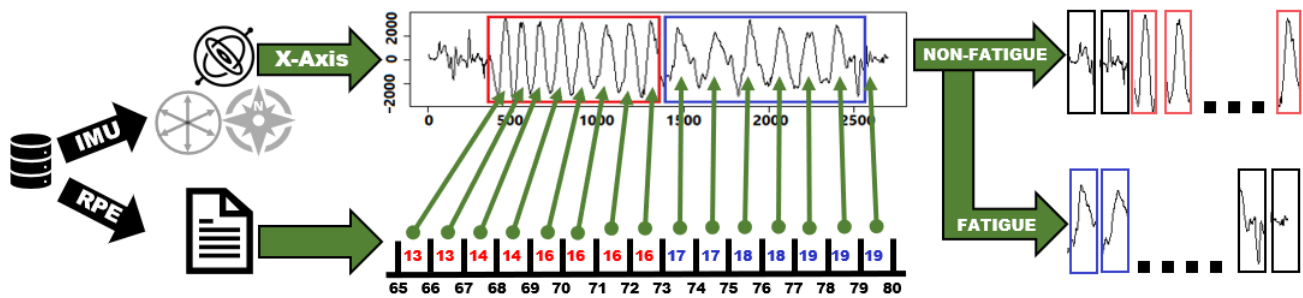
Figure 4. An example of extracting and labeling repetitions of the fifth set from the gyroscope's x-axis.

However, the process is not as exhausting as it may seem per volunteer because we use a synchronized IMU. Therefore, we only need to fully extract and label all 5 sets recorded by one axis per volunteer, e.g., gyroscope's x-axis. Then, we use the same timestamps from the gyroscope's x-axis to extract and label repetitions for all the remaining signals (other axes of gyroscope and accelerometer). After we processed all data from the volunteers, our dataset consisted of 3750 repetitions recorded from 6 time-series signals. We compute three additional signals, which are total acceleration, exerted force, and acc–gyro data fusion, as the following:

- Total acceleration: This is the vector sum of the tangential and centripetal accelerations, which makes it a place-independent signal. Hence, total acceleration does not rely on the exact attachment of the accelerometer because it combines x, y, and z acceleration signals at time t_i to compute a total acceleration, defined as: $\sqrt{a_{x_i}^2 + a_{y_i}^2 + a_{z_i}^2}$.

- Exerted force: $F_{exerted} = m \times a$ is the exerted force by the volunteer to lift the dumbbell. $F_{exerted}$ is calculated by multiplying the mass m of the lifted dumbbell by acceleration a.

- Acc–gyro data fusion (complementary filter): A complementary filter is often used to detect human body movement patterns by combining the gyroscope and the accelerometer [66,67]. Gyroscope's data are used for precision because it is not vulnerable to external forces, while the accelerometer's data are used for long-term tracking as it does not drift. We use the Kalman filter algorithm to estimate roll, pitch, and yaw angles [68]. However, we use the yaw angle because it indicates the sideways vibration for the volunteer's hand during the extension and flexion of the bicep. Previous studies show fatigue may cause a temporary movement disorder, such as skeletal muscles vibration, which indicates fatigue backlogs and increases the vibration angle [69–72]. In the filter's simplest form, the equation is defined as: $angle = 0.98 \times (angle + gyro \times dt) + 0.02 \times acc$.

2.3. Feature Extraction

Feature extraction is a crucial component of HAR systems because it establishes the most significant parameters to identify or predict human body movements. In addition, feature extraction reduces the data dimensionality while preserving the relevant characteristics of the signal. In this work, we compute a total of eleven hand-crafted features, as shown in Table 2. Eight of the selected features are proven accurate in previous works, especially in the general classification of human activities [15,33,34,73–75]. These features include min, max, mean, median, SD, variance, kurtosis, and RMS. Besides the eight features mentioned before, we also select three other features often associated with fatigue for better performance: skewness, IoP, and MSP [76–79]. A previous study suggests considering the skewness of the data when detecting fatigue in repetitive muscle movements such as bicep curls [55]. We select skewness as a fatigue feature because, during the repetitions' extraction and labeling process, we observed the following: (1) Non-fatigue repetitions are relatively symmetrical during the repetitions' extraction and labeling process. (2) In

contrast, fatigue repetitions are often positively skewed. Another work shows that fatigue often occurs in later sets, which increases the time to complete repetitions of bicep curls while decreasing the force exerted by the muscles [45]. This is observable through the increments of intervals between peaks, e.g., IoP, and decrements of peaks' amplitudes, e.g., MSP. For each volunteer, we extract the eleven features on all repetitions, across all nine signals, including the two 3D signals (x, y, z) from the accelerometer and the gyroscope, total acceleration, exerted Force, and acc–gyro signal fusion.

Table 2. Eleven hand-crafted features: eight HAR-related features and three fatigue-related features.

	Feature	Formula
Centralized	Minimum	$min = min_{i=1,...N}(x_i)$
	Maximum	$max = max_{i=1,...N}(x_i)$
	Mean	$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$
	Median	$M = \begin{cases} x_{\frac{N+1}{2}}, \text{ N odd} \\ \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}), \text{N even} \end{cases}$
	Standard Deviation (SD)	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N}(x_i - \overline{x})^2}$
	Variance	$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n}(x_i - \overline{x})^2$
	Kurtosis	$K = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \overline{x})^4}{\sigma^4}$
	Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N}(x_- \overline{x})^2}$
Fatigue	Skewness	$Sk = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \overline{x})^3}{\sigma^3}$
	Interval of Peaks (IoP)	$IoP = T_p - T_{p-1} : p = 2,...N$
	Mean Slope between Peaks (MSP)	$MSP = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{p_j - p_i}{T_{p_j} - T_{p_i}}$

2.4. Extracting Similarities

To visualize the concept of our work, let us assume that a test subject is selected from a diverse population P of size n, as shown in Figure 5. Each member of the population reports their physical traits along with bicep concentration curl data signals. Meanwhile, the test subject provides only partial data, often one set of repetitions, of their bicep concentration curl data signals along with their physical traits. We measure the similarities between the test subject and members of the population so that the data from whom the test subject is similar gain more weight while training the model. We are keen to utilize two types of similarities, physical similarity and signal similarity, because previous studies have reported gains in performance when harnessing those similarities to weight data from the crowd [28,31].

We believe that combining both physical and signal similarities may further improve the personalized models' performance. Therefore, in the second research question (RQ2), we measure and compare the performance between the personalized models trained using the weighted data and the cross-subject models. Furthermore, it comes to our minds that if we already possess and use a part of the test subject's data to measure the similarities between the test subject and the crowd, then we may let the personalized models consume it in training to improve their learning. Therefore, in the third research question (RQ3), we also decided to let subject-specific models consume the same part of the test subject's data; then, we compare the performances of personalized and subject-specific models. Moreover, we allow the subject-specific models to consume more of the test subject's data if needed

until it can reach the same performance of personalized models so that we quantify the amount of spared data by using personalized models.

Figure 5. Visualization of the concept of personalizing general model using crowd-sourced wearables' data.

2.4.1. Measuring Physical Similarity

Physical characteristics of people (e.g., age, weight, height, or BMI) vary from one person to another within a large population. Such differences can affect the way people move and perform physical activities. We believe that a user with different physical traits, e.g., age and BMI, may show signs of fatigue differently. At the same time, we expect groups of people who share similar physical traits to show similar signs of fatigue [80,81]. For example, let us capture the signs of fatigue using the three fatigue-related features, skewness, IoP, and MSP, and plot the principal component analysis (PCA). In Figure 6a, we often observe that individuals within specific limits of BMI values tend to share similar signs of fatigue. Moreover, Figure 6b shows similar observations where we use age instead of BMI. We can observe that individuals of certain ages tend to share similar signs of fatigue. This strengthens our hypothesis that if we construct the training data from individuals within the population who are more similar, it may reduce the inter-subject data variability and hence improve the performance of the fatigue detection model.

(**a**)   (**b**)

Figure 6. PCA plots showing signs of fatigue captured by the three fatigue-related features and BMI/age. (**a**) BMI perspective. (**b**) Age perspective.

To compute the physical similarity value between a pair of users, we employ four types of physical traits: age, height, weight, and BMI. To limit the widespread of the values, due to subjects' variations, we apply min–max normalization to each physical trait, on training data, to normalize each trait between 0 and 1. We combine these four traits per user to form a dedicated physical vector V^{Phy} = {age, height, weight, BMI} representing their physical traits separately. We measure the distance d^{Phy} between the physical traits V^{Phy} of two users (q, p) based on the Manhattan distance, as shown in Equation (1). Previous works show that Manhattan distance is preferable to Euclidean for high dimensional data and if the dimensions are not comparable [82–84]. The physical similarity between users (q, p) is based on the universal law of generalization proposed in previous works [28,31,85,86],

where distance and perceived similarity are related via an exponential function, as shown in Equation (2):

$$d^{Phy}(q,p) = \sum_{k=1}^{4} |V_{q_k}^{Phy} - V_{p_k}^{Phy}| \tag{1}$$

$$sim^{Phy}(q,p) = \frac{1}{e^{\gamma d^{Phy}(q,p)}} \tag{2}$$

where γ is an empirically determined scaling parameter that affects the shape of the exponential function. For example, $\lim_{\gamma \to \infty} sim^{Phy}(q,p) = 0$, which indicates that as γ approaches infinity, the physical similarity approaches zero, causing more segregation between users. This can be a double-edged sword because as we segregate dissimilar users from each other, we may increase the segregation between similar users unintentionally. On the other hand, $\lim_{\gamma \to 0} sim^{Phy}(q,p) = 1$, which indicates that as gamma approaches zero, the physical similarity approaches one, implying that all subjects show similar signs of fatigue; in other words, the changes in their data patterns are similar. Again, this is a double-edged scenario where we may unintentionally pull dissimilar users near to the similar users. Therefore, further investigation is required to estimate the optimal value of γ.

2.4.2. Measuring Signal Similarity

In the context of signal similarities, we use one set of repetitions, approximately 20% of the subject's data needed for the subject-specific models. We believe that users within the same population may show similar signs of fatigue, leading to similar changes in data patterns while performing the exercise. To compute the signal similarity value between a pair of users, we employ the 11 extracted features in Table 2 to form a dedicated signal vector $V^{Sig} = \{min, max, \dots, MSP\}$ for each user. We measure the distance d^{Sig} between the signal traits V^{Sig} of two users (q,p) based on the Manhattan distance for all repetitions $l = \{1, 2, \dots, L\}$, as shown in Equation (3).

$$d^{Sig}(q,p) = \sum_{k=1}^{11} \sum_{l=1}^{L} |V_{q_{(k,l)}}^{Sig} - V_{p_{(k,l)}}^{Sig}| \tag{3}$$

$$sim^{Sig}(q,p) = \frac{1}{e^{\gamma d^{Sig}(q,p)}} \tag{4}$$

The signal similarity between users (q,p) is based on the distance between their vectors, as shown in Equation (4).

2.4.3. Measuring Total Similarity

We measure the total similarity sim^{Total} between two users (q,p) by summing their weighted physical $sim^{Phy}(q,p)$ and signal $sim^{Sig}(q,p)$ similarities, as shown in Equation (5).

$$sim^{Total}(q,p) = \alpha \times sim^{Phy}(q,p) + \beta \times sim^{Sig}(q,p) \tag{5}$$

where $\alpha + \beta = 1$. If α is greater than β, the physical similarity will contribute more than signal similarity in determining the total similarity value. On the other hand, if β is greater than α, the signal similarity will be the one that dominates the total similarity value. Therefore, in the first research question (RQ1), we further investigate the impact of (α, β) values on the performance of the personalized models. Moreover, we examine (γ) values to achieve the highest performance possible.

3. Experiments and Results

In this section, we describe the setup to evaluate the personalization approach in boosting the performance of cross-subject models. We start by listing the three research

questions and the selected models for the experiment setup. Then, we identify the possible values for the parameters (α, β, γ) in Equations (2), (4), and (5).

- RQ1: What is the impact of the physical and signal parameters on the performance of the personalized biceps fatigue detection models?
- RQ2: Can the personalization approach improve the performance of cross-subject models in detecting biceps muscle fatigue?
- RQ3: Can the personalization approach reduce the consumption of the test subject's data in comparison to subject-specific models?

We employ two models that utilize weighted data in the training to improve classification performance. Several works in the literature suggest using AdaBoost, a statistical boosting technique used as an ensemble method to reduce error in generalized models at the cost of weighing the training data [87–89]. Therefore, we use AdaBoost to take advantage of weighting the data, as mentioned in Section 2.4, while training these models. We select two models that have been previously used in HAR-related applications. The first model is the Decision Tree (DT); this model was used to count and classify ambulatory activities using eight plantar pressure sensors within smart shoes in a previous study [90]. The second model is the Artificial Neural Networks (ANN); this model was used to detect and count repetitions for complex physical exercises [91]. Specifically, we use Adaboost-backpropagation neural network, which is used in HAR-related works [92,93]. There are two reasons for choosing DT and ANN among several other machine learning algorithms. (1) The current work aims to mitigate the hindering effect of subject data variability on cross-subject models, which we previously encountered in our prior work [45]. Therefore, we used the same models as in our previous work to fairly examine whether our personalization approach can minimize such a problem. (2) We wanted to examine a previous study finding that AdaBoost often achieves higher improvement on weak classifiers such as DT than stronger ones such as ANN [92].

In this work, we construct the training dataset based on the similarity score between the test subject and the crowd. The data from whom the test subject has the highest similarity score are used in validation/ tuning the models, while the remaining data are used to train the models. On the other hand, the test dataset is unseen data collected from the test subject and used only to assess the performance of the models.

3.1. Examining the Hyper-Parameters in the Personalized Biceps Fatigue Detection Model

In this section, we state our motivation and approach for RQ1. Our motivation for RQ1 is that we believe the performance of users similarity-based models, such as those driven from the personalization approach, may degrade if the inadequate parameters are selected. In other words, valuable data from the crowd, e.g., similar users, may be discarded due to an unintended preference for the physical similarity over signal similarity and vice versa. Previous work shows that finding a balance between the extracted similarities is important to improve the accuracy of the models constantly [94].

Our approach to address RQ1 is that we examine the possible values for the parameters (α, β, γ) in Equations (2), (4) and (5) so that we observe the impact of these parameters on the models' performance. We start with Equations (2) and (4) to examine γ, which has an arbitrary value between $(0, \infty)$. To observe the impact of the different γ values on the models' performance, we employ each γ value to run two pairs of models: AdaBoost-DT and AdaBoost-ANN. The first pair is physical similarity-based models of AdaBoost-DT and AdaBoost-ANN, while the second pair is signal similarity-based models of AdaBoost-DT and AdaBoost-ANN. Then, we compute the changes in these models' performance as the value of γ increases and select the gamma value that corresponds to the best performance. In Equation (5), we examine the physical α and signal β parameters while satisfying the condition $\alpha + \beta = 1$. We experiment with different α and β values and observe the effect by running two similarity-based models (AdaBoost-DT, AdaBoost-ANN). Then, we compute the models' performance as the values of α and β change; then, we select the (α, β) values that correspond to the best performance.

3.1.1. RQ1: What Is the Impact of the Physical and Signal Parameters on the Performance of the Personalized Biceps Fatigue Detection Models?

Figure 7 shows the average changes in both models' performance as the value of γ increases. The performance in this context is measured using accuracy. We use the accuracy at $\gamma = 0$ as the reference point to measure the changes in accuracy ($\Delta Accuracy$) as the γ value increases. We observe that $\Delta Accuracy$ increases as the γ value increases until both reach maximum values of 3.83 and 14 at the dashed line, respectively. Then, the changes in accuracy start to decline as the γ value continues to increase. Therefore, we select the $\gamma = 14$ to let the models perform at maximum accuracy.
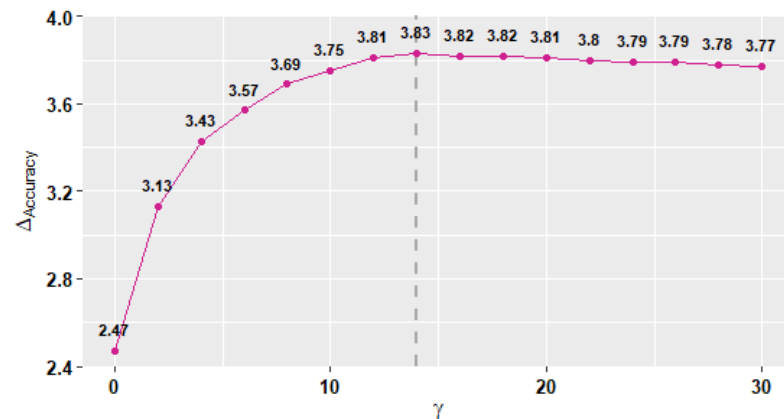
Figure 7. The average changes in both models' accuracy as the value of γ increases.

Figure 8 shows the impact of the physical α and signal β parameters on models' performance and there are three important findings in the figure. In the first finding, at $\alpha = 0$ and $\beta = 1$, the models (AdaBoost-ANN, AdaBoost-DT) solely depend on the signal similarity, leading the training dataset, for these models, to be selected from the crowd with whom the test subject's signal is similar while discarding the physical similarity. In this case, the two models (AdaBoost-ANN, AdaBoost-DT) achieve accuracy of 82.13% and 62.49%, respectively. In the second finding, at $\alpha = 1$ and $\beta = 0$, the models (AdaBoost-ANN, AdaBoost-DT) solely depend on the physical similarity, leading the training dataset, for these models, to be selected from the crowd with whom the test subject's physical traits is similar while discarding the signal similarity. In this case, the accuracy for the two models (AdaBoost-ANN, AdaBoost-DT) drops to 81.07% and 62.52%, respectively. In the third finding, at $\alpha = [0.25, 0.50]$ and $\beta = (1 - \alpha)$, the models (AdaBoost-ANN, AdaBoost-DT) depend on both physical and signal similarities; however, the training dataset, for these models, is selected from the crowd with whom the test subject's is similar while prioritizing those with the highest signal similarity. In this case, we observe that the accuracy for the two models (AdaBoost-ANN, AdaBoost-DT) rises to reach 84.54% and 65.88%, respectively.
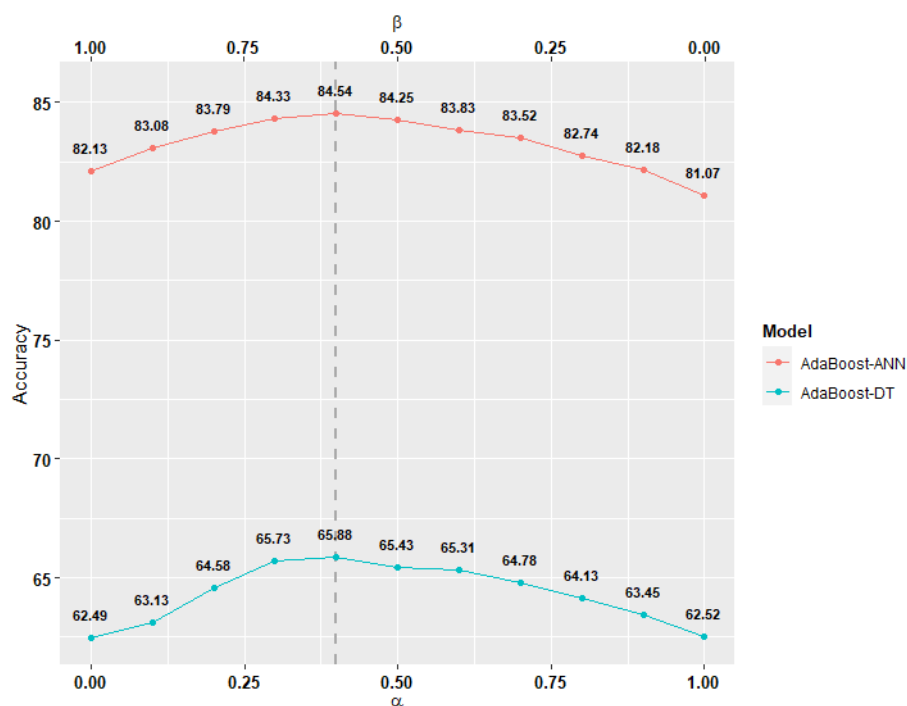
Figure 8. The average models' accuracy as the values of α and β change.

3.1.2. RQ1 Conclusion

Our findings show that both physical and signal similarities are important. Moreover, the models' performance reaches its peak when $\gamma = 14$. The best selected values for $\alpha = 0.4$ and $\beta = 0.6$ which we use for the rest of the evaluations.

3.2. Evaluating the Performance of Personalized Models

In this section, we state our motivation and approach for RQ2. Our motivation for RQ2 is that cross-subject models may seem preferable when it comes to a large number of users. However, a common trade-off for having cross-subject models is accuracy loss, especially for users with particular activity patterns. In other words, users who do not share enough similarities with the crowd may look as outliers where the cross-subject models are less accurate to detect their biceps muscle fatigue. We believe that adding weight to user's data from whom the test subject is similar can improve model accuracy, including marginal users. Results of a previous study show that the personalization of cross-subject models constantly improves their accuracy compared with the standard cross-subject models [27].

$$Accuracy = \frac{True(Fatigue + NonFatigue)}{True(Fatigue + NonFatigue) + False(Fatigue + NonFatigue)} \quad (6)$$

$$Precision = \frac{True(Fatigue)}{True(Fatigue) + False(Fatigue)} \quad (7)$$

$$Recall = \frac{True(Fatigue)}{True(Fatigue) + False(NonFatigue)} \quad (8)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (9)$$

Our approach to address RQ2 is that we use the 11 hand-crafted features along with the 2 similarity-based models (AdaBoost-DT, AdaBoost-ANN). We use these models to predict the Borg rating for each repetition, detecting whether a repetition contains fatigue or not. We run two experiments. In the first experiment, we set ($\alpha = 0, \beta = 0$) in both models to mimic the standard cross-subject models. This means that the training dataset for these

models is collected without considering any type of similarity between the crowd and the test subject. For this experiment, we use leave-one-out cross-validation (LOOCV), which is a K-fold cross-validation with K equal to the number of volunteers ($K = 25$). In the second experiment, we set (α, β) to optimal values as identified in Section 3.1.1, leading the training dataset, for these models, to be selected based on physical and signal similarities, in addition to prioritizing data coming from users of highest signal similarity in the crowd. For each experiment, we calculate the accuracy using the confusion matrix shown in Table 3, where non-fatigue repetition represents a Borg score from 6 to 16, and fatigue status represents a Borg score from 17 to 20. We calculate the accuracy using Equation (6), precision using Equation (7), recall using Equation (8), and F1 using Equation (9).

Table 3. Fatigue detection confusion matrix.

		Actual	
		Fatigue $\in$ [17,20]	**Non-Fatigue $\in$ [6,16]**
Predict	**Fatigue $\in$ [17,20]** **Non-Fatigue $\in$ [6,16]**	TRUE Fatigue FALSE Non-Fatigue	FALSE Fatigue TRUE Non-Fatigue

3.2.1. RQ2: Can the Personalization Approach Improve the Performance of Cross-Subject Models in Detecting Biceps Muscle Fatigue?

It is important to mention that we use one set of 15 repetitions approximately from the test subject's data during the personalization of DT and ANN to measure the signal similarity between the test subject and individuals in the crowd. However, we do not include these 15 repetitions nor any data from the test subject in the training set. Our findings show that the personalization approach improves the accuracies for the models by 5.89% (DT) and by 3.38% (ANN), as shown in Table 4. The accuracy improved after prioritizing training data from the crowd based on the total similarity score, where individuals with high scores contribute more to the models' training dataset. Moreover, we observe other improvements in terms of precision, recall, and F1-measure across the models. The results show that the personalization improves the DT model in terms of precision by (1.20%), recall by (4.51%), and F1-measure to (2.81%). On the other hand, the personalization improves the ANN model in terms of precision by (6.96%), recall by (4.55%), and F1-measure to (5.82%). Moreover, we observe that the standard cross-subject ANN model outperforms both DT models, which is expected in fatigue detection wherein ANN models often perform better than other cross-subject models [95].

Table 4. Average precision, recall, and accuracy, with a CI of 95%, for detecting fatigue in biceps repetitions before and after the personalization of cross-subject models.

	Models					
		DT			**ANN**	
	Cross-Subject	**Personalized**	**Δ**	**Cross-Subject**	**Personalized**	**Δ**
Precision	60.57% $\pm$ 0.66	61.77% $\pm$ 0.58	1.20%	73.29% $\pm$ 0.43	80.25% $\pm$ 0.47	6.96%
Recall	61.32% $\pm$ 0.53	65.83% $\pm$ 0.49	4.51%	78.53% $\pm$ 0.39	83.08% $\pm$ 0.43	4.55%
Accuracy	60.08% $\pm$ 0.49	65.97% $\pm$ 0.67	5.89%	82.41% $\pm$ 0.58	85.79% $\pm$ 0.48	3.38%
F1	60.94% $\pm$ 0.59	63.75% $\pm$ 0.53	2.81%	75.82% $\pm$ 0.41	81.64% $\pm$ 0.45	5.82%

3.2.2. RQ2 Conclusion

Overall, our findings indicate that the personalization approach improves both models in terms of performance. For the DT model, the personalization improves its F1-measure from (60.94%) to (63.75%), while for the ANN model, the personalization improves its F1-measure from (75.82%) to (81.64%).

3.3. Examining the Consumption of the Test Subject's Data in the Personalization Approach

In this section, we state our motivation and approach for RQ3. Our motivation for RQ3 is that subject-specific models are known for their high performance and demand of the test subject data. On the other hand, the cross-subject models are known for their relatively lower performance and no test subject data demand. We believe that the personalized approach can combine the best aspects of these two models. In other words, the personalization approach can improve the performance of the cross-subject models while consuming less test subject data than the subject-specific models. A previous work shows that adding a small amount of the test subject's data to the training dataset for personalized models helps to improve the performance further closer to the subject-specific models [36].

Our approach to address RQ3 is that we utilized the 11 hand-crafted features and 4 models: subject-specific (DT, ANN) models and personalization (AdaBoost-DT, AdaBoost-ANN) models. We use these models to predict the Borg rating for each repetition to determine whether it is fatigue repetition or not. Similar to RQ2, we set (α, β) to optimal values so that the training dataset for these models is selected based on physical and signal similarities while prioritizing the signal similarity selection. Our experiment consists of seven runs where we incrementally add 10% of the test subject's data to the training set after each run. This means, there is 0% of the test subject's data added to the training dataset at the 1st run, while at the 7th run, there is 60% of the test subject's data added to the training dataset. We stop at 60% of the test subject's data to prevent overfitting the personalized model; otherwise, there will not be much of a difference between the subject-specific and the personalization approaches. Moreover, this allows us to measure the amount of the test subject's data needed to improve the personalization models' performance closer to the subject-specific models. Or, in other words, how little data we need from the test subject if we use the personalized models instead of subject-specific ones while keeping the accuracy relativity high. For each run, we calculate the accuracy using Equation (6) and accuracy gain ratio (AGR) using Equation (10).

$$AGR = \frac{\Delta accuracy}{\text{Test subject's data (repetitions)}} \tag{10}$$

3.3.1. RQ3 Results: Can the Personalization Approach Reduce the Consumption of the Test Subject's Data?

Our findings show that the more the test subject's data are added to the training set, the higher the accuracy of the subject-specific and personalized models. Table 5 shows that the subject-specific DT model achieves an accuracy of 78.90% after consuming 40% of the test subject's data. On the other hand, the personalization of the DT model achieves an accuracy of 76.08% after consuming 20% of the test subject's data while compensating the rest of the training data from the similar users in the crowd. In other words, the subject-specific DT model requires twice the amount of test subject data, at 40%, to achieve similar accuracy to the personalized DT at 20% of the test subject's data consumption with taking into consideration that the personalized DT model compensates the rest of the training data from the crowd. Moreover, our findings show that with 20% of the test subject's data, the personalized DT model reaches the lowest accuracy gain ratio of 0.36% per test subject's repetition while maintaining the highest accuracy gain of 5.44%.

Furthermore, the subject-specific ANN model achieves an accuracy of 92.99% after consuming 60% of the test subject's data. On the other hand, the personalization of the ANN model achieves an accuracy of 92.74% after consuming 20% of the test subject's data while compensating the rest of the training data from the similar users in the crowd. In other words, the subject-specific ANN model requires triple the amount of test subject's data, at 60%, to achieve similar accuracy to the personalized ANN at 20% of the test subject's data consumption with taking into consideration that the personalized ANN model compensates the rest of the training data from the crowd. Moreover, our findings show that with 20% of the test subject's data, the personalized ANN model reaches the

lowest accuracy gain ratio of ≈0.36% per test subject's repetition while maintaining the highest accuracy gain of 5.37%.

Table 5. The accuracy averages for the subject-specific and personalized models after adding 10% of the test subject's data to the training set in each run incrementally. We include a version of this table with the confidence intervals in the Appendix A.

			Number of Biceps Repetitions Collected from the Test Subject (% of Used Test's Data)						
			0 (0%)	8 (10%)	15 (20%)	23 (30%)	30 (40%)	38 (50%)	45 (60%)
Models	**DT**	**Subject-specific**	Accuracy	15.34%	41.30%	58.40%	68.60%	78.90%	82.20%	84.03%
			$\Delta accuracy$	-	25.96%	17.10%	10.20%	10.30%	3.30%	1.83%
			AGR	-	3.25%	1.14%	0.44%	0.34%	0.09%	0.04%
		Personalized	Accuracy	65.88%	70.64%	76.08%	77.77%	79.15%	79.55%	79.91%
			$\Delta accuracy$	-	4.76%	5.44%	1.69%	1.38%	0.40%	0.36%
			AGR	-	0.60%	0.36%	0.07%	0.05%	0.01%	0.01%
	ANN	**Subject-specific**	Accuracy	55.23%	62.48%	74.68%	82.95%	87.37%	90.45%	92.99%
			$\Delta accuracy$	-	7.25%	12.20%	8.27%	4.42%	3.08%	2.54%
			AGR	-	0.91%	0.81%	0.36%	0.15%	0.08%	0.06%
		Personalized	Accuracy	84.54%	87.37%	92.74%	93.56%	93.88%	94.25%	94.78%
			$\Delta accuracy$	-	2.83%	5.37%	0.82%	0.32%	0.37%	0.53%
			AGR	-	0.35%	0.36%	0.04%	0.01%	0.01%	0.01%

3.3.2. RQ3 Conclusion

Our findings show that the personalization approach may reduce the test subject's data consumption by 33.3% up to 50.0% while reducing the accuracy gap compared to the subject-specific models.

4. Discussion

This section discusses our findings, points out work limitations, and proposes future works.

4.1. Findings Discussion

In RQ1, we observe that $\Delta Accuracy$ increases following a γ increase until it reaches a maximum of 3.83 at $\gamma = 14$. Then, the accuracy starts to drop slightly with higher values of γ. Previous studies report similar behavior for the gamma parameter in their results sections [28,31]. They observe their models' accuracy increases with an increase in γ value until γ reaches an optimal point, where their models then start losing accuracy. Although gamma's behavior seems similar, the γ values are different and depend on the dataset. The reason behind gamma's behavior resides in Equations (2) and (4) where we find that the physical and signal similarities approach zero as $\gamma \rightarrow \infty$. This means, if we keep increasing the gamma values, we will push the test subject further away from the crowd. In other words, we, unintentionally, decrease the possibility of finding similar users in the crowds, resulting in fewer similar data points, and hence smaller training data. On the other hand, when $\gamma \rightarrow 0$, the physical and signal similarities approach 1. This means, if we keep decreasing the gamma values, we will push the test subject closer toward the crowd, increasing the possibility of finding similar users. However, this can increase the risk of including low-quality data points from similar users with low ranks, which is usually the case in a cross-subject model; therefore, the accuracy often drops.

In RQ2, our findings show improvements in accuracy for both personalized models, AdaBoost-DT and AdaBoost-ANN, compared to the standard cross-subject ones, by 5.89% and 3.38%, respectively. Such improvements occur because the training datasets for the personalized models are selected from users who's physical and signal traits are similar to the test subject. A previous study reports similar findings to ours, indicating that personalized

models often perform +3% better than standard cross-subject models [27]. However, their proposed personalized model averages 0.78 for F1-score, while our AdaBoost-ANN model performs 3.64% better with an average of 81.64% $\pm$ 0.45 for F1-score. Overall, while this has been shown in previous works, our results help consolidate the benefits on relying on similarity as a method for boosting the performance of cross-subject models.

In RQ3, our finding indicates that the personalized models, AdaBoost-DT and AdaBoost-ANN, achieve comparable performance to subject-specific models while consuming 50.0% and 66.77% less test subject data. This is an important finding to motivate approaches that rely less on the data of the subject, particularly in cases where the test subject's data are difficult to obtain or very limited. A previous study utilized a personalization approach to cut down the cost of data labeling by up to 90% for new users [37]. The study reports model accuracy between 77.7% and 83.4%. In contrast, we can observe that our personalized models, AdaBoost-DT and AdaBoost-ANN, achieve closer or higher accuracies at 76.08% $\pm$ 0.71 and 92.74% $\pm$ 0.49 at similar rates of 20% test data consumption, respectively, as shown in Table 6. This table shows the accuracy achieved by fatigue detection models including the cross-subject, subject-specific, and personalized models. As an implication, our findings suggest that personalized models are an effective approach to reduce data dependency—when data on the target subject is scarce—without severely compromising the model's performance.

Table 6. Percent accuracy achieved on, with a CI of 95%, the cross-subject, subject-specific, and personalization models.

		(% of Used Test's Data)	Accuracy	ΔAccuracy		
				Cross-Subject	**Personalized**	**Subject-Specific**
Models	**DT**	Cross-Subject (0%)	60.08% $\pm$ 0.49	-	−16.00%	−28.67%
		Personalized (20%)	76.08% $\pm$ 0.71	16.00%	-	−12.67%
		Subject-Specific (100%)	88.75% $\pm$ 0.59	28.67%	12.67%	-
	ANN	Cross-Subject (0%)	82.41% $\pm$ 0.58	-	−10.33%	-16.89%
		Personalized (20%)	92.74% $\pm$ 0.49	10.33%	-	−6.56%
		Subject-Specific (100%)	99.30% $\pm$ 0.37	16.89%	6.56%	-

Moreover, we can observe that both of the personalization models achieve higher accuracies compared to the cross-subject models. However, we find that the personalization ANN models achieve lesser accuracy improvement than the personalization DT models. This agrees with previous work that shows that AdaBoost usually achieves higher improvement results on weak classifiers such as DT than stronger ones such as ANN [92].

4.2. Work Limitations

The first limitation of our work is the data size, which may affect the external validity of our study. While some HAR studies have opted to use public datasets, datasets with fatigue data are not common nor often available to the public [31,96]. Since we have to collect our fatigue data during the COVID-19 pandemic, it has been a daunting task due to social distancing and restrictive measures. Although our dataset may look small in size, we believe it is suitable for our research under such circumstances as other studies also collected their dataset with similar sizes to ours [97,98]. We agree that a bigger dataset is beneficial to our work, but we believe our experiments/approach can generate similar performance approximately.

The second limitation of our work is the reliance on the Apple Watch Series which uses the photoplethysmography (PPG) sensor to measure participants' heart rate during the exercise. Although Apple Watch can provide the most accurate readings amongst the optical wrist wearables [99], previous works show that PPG often suffers from inaccuracies. This means our results may be indirectly impacted [100,101]; however, we believe such technology does not compromise our findings, especially in real-life applications. Previous

works show that PPG achieves clinically acceptable accuracy and might be considered safe for rehabilitation training programs [102,103].

The third limitation of our work is the use of the Borg scale and the dumbbell weight. Although the Borg scale is often used in sports science, some studies are often cautious about its implications [104,105]. Using subjective measures such as the Borg scale to report RPE may introduce a dependency between the correctness of selected Borg rating and participants' awareness. Therefore, we introduce the concept of the Borg scale to the participants in advance to avoid misevaluating their perceived exertion rate. On the other hand, dumbbell weight can directly vary the data points collected during the exercise because of each participant's physical capacity or strength. A previous study shows lightweight dumbbells lead to a long recording session with many similar data points until participants reach fatigue [54]. In contrast, heavyweight dumbbells lead to shorter recording sessions with fewer data entries, which do not capture kinetic changes clearly throughout the exercise because participants reach fatigue quickly. Although we use a 4.5 kg weight dumbbell as recommended by previous studies, we believe having dumbbell weights will provide us with more information and different patterns of biceps muscle fatigue [51–53].

4.3. Future Works

The current work aims to mitigate the hindering effect of subject data variability from our prior works based on the Borg scale [45,55]. To achieve fairness in evaluation and continuity of our prior works, we use the Borg scale in our personalization approach. However, previous studies show the advantages of Electromyography (EMG) sensors in providing detailed information about muscle conditions during incremental exercises [97]. Moreover, a recent study that utilizes the EMG sensors shows an average accuracy of 97.8% in muscle fatigue detection during incremental exercises [106]. We believe utilizing such sensors can improve the performance of our personalization approach. However, collecting and processing EMG sensors data from bicep muscles will require effort, coding, and time; therefore, we opt to select EMG sensors in our future work.

Another future work that we are interested in is examining the personalization approach using deep learning such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNN). A recent study shows a significant performance improvement (from 89.83% to 96.62%) after utilizing deep learning in HAR systems [107]. We opt to use DT and ANN models in the current study to examine whether our personalization approach can mitigate the hindering effect of subject data variability encountered in our previous works [45,55]. So, since our findings seem encouraging, we have decided to use CNN in our future work to examine our personalization approach further.

5. Conclusions

This work aims to mitigate the hindering effect of subject data variability in fatigue detection. We propose the personalization approach to utilize data from the crowd based on the total similarity score between the test subject and the crowd. Our dataset consists of 3750 concentration curl repetitions from 25 volunteers with ages and BMI ranging between 20–46 and 24–46, respectively. We compute the total similarity score between each individual in the crowd and the test subject based on the physical and signal similarity scores. Then, we extract a weighted dataset to train our models. Our findings show that the AdaBoost-DT model outperforms the cross-subject-DT model by 5.89%, while the AdaBoost-ANN model outperforms the cross-subject-ANN model by 3.38%. On the other hand, at 50.0% less of the test subject's data consumption, our AdaBoost-DT model outperforms the subject-specific-DT model by 16%, while the AdaBoost-ANN model outperforms the subject-specific-ANN model by 10.33%. Our findings indicate that crowd data are usable to build personalized bicep fatigue detection models to prevent athletes from fatigue-induced injuries. Moreover, our personalization approach benefits real-life applications when the data from the test subject is unavailable or insufficient. We believe that our work is useful and represents a

solid start for moving into real-world applications for detecting the fatigue level in bicep muscles using wearables' data from the crowd.

Appendix A

Table A1. The accuracy averages for the subject-specific and personalized models after adding 10% of the test subject's data to the training set in each run incrementally.

Models				Number of Biceps Repetitions Collected from the Test Subject (% of Used Test's Data)						
				0 (0%)	8 (10%)	15 (20%)	23 (30%)	30 (40%)	38 (50%)	45 (60%)
	DT	Subject-specific	Accuracy	15.34% ± 0.83	41.30% ± 0.78	58.40% ± 0.71	68.60% ± 0.63	78.90% ± 0.57	82.20% ± 0.48	84.03% ± 0.41
			$\Delta accuracy$	-	25.96%	17.10%	10.20%	10.30%	3.30%	1.83%
			AGR	-	3.25%	1.14%	0.44%	0.34%	0.09%	0.04%
		Personalized	Accuracy	65.88% ± 0.65	70.64% ± 0.68	76.08% ± 0.71	77.77% ± 0.66	79.15% ± 0.59	79.55% ± 0.51	79.91% ± 0.47
			$\Delta accuracy$	-	4.76%	5.44%	1.69%	1.38%	0.40%	0.36%
			AGR	-	0.60%	0.36%	0.07%	0.05%	0.01%	0.01%
	ANN	Subject-specific	Accuracy	55.23% ± 0.79	62.48% ± 0.72	74.68% ± 0.64	82.95% ± 0.58	87.37% ± 0.45	90.45% ± 0.47	92.99%
			$\Delta accuracy$	-	7.25%	12.20%	8.27%	4.42%	3.08%	2.54%
			AGR	-	0.91%	0.81%	0.36%	0.15%	0.08%	0.06% ± 0.33
		Personalized	Accuracy	84.54% ± 0.63	87.37% ± 0.55	92.74% ± 0.49	93.56% ± 0.42	93.88% ± 0.48	94.25% ± 0.45	94.78% ± 0.39
			$\Delta accuracy$	-	2.83%	5.37%	0.82%	0.32%	0.37%	0.53%
			AGR	-	0.35%	0.36%	0.04%	0.01%	0.01%	0.01%

References

1. Enoka, R.M.; Duchateau, J. Muscle fatigue: What, why and how it influences muscle function. *J. Physiol.* **2008**, *586*, 11–23. [CrossRef]
2. Opar, D.; Williams, M.; Shield, A. Hamstring Strain Injuries Factors that Lead to Injury and Re-Injury. *Sport. Med.* **2012**, *42*, 209–226. [CrossRef]
3. Mueller-Wohlfahrt, H.W.; Haensel, L.; Mithoefer, K.; Ekstrand, J.; English, B.; McNally, S.; Orchard, J.; van Dijk, C.N.; Kerkhoffs, G.M.; Schamasch, P.; et al. Terminology and classification of muscle injuries in sport: The Munich consensus statement. *Br. J. Sport. Med.* **2013**, *47*, 342–350. [CrossRef]
4. Kellmann, M. Preventing overtraining in athletes in high-intensity sports and stress/recovery monitoring. *Scand. J. Med. Sci. Sports* **2010**, *20*, 95–102. [CrossRef]
5. Thalman, C.M.; Lam, Q.P.; Nguyen, P.H.; Sridar, S.; Polygerinos, P. A Novel Soft Elbow Exosuit to Supplement Bicep Lifting Capacity. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 6965–6971.
6. Stoudemire, N.M.; Wideman, L.; Pass, K.A.; Mcginnes, C.L.; Gaesser, G.A.; Weltman, A. The validity of regulating blood lactate concentration during running by ratings of perceived exertion. *Med. Sci. Sports Exerc.* **1996**, *28*, 490–495. [CrossRef]
7. Crewe, H.; Tucker, R.; Noakes, T.D. The rate of increase in rating of perceived exertion predicts the duration of exercise to fatigue at a fixed power output in different environmental conditions. *Eur. J. Appl. Physiol.* **2008**, *103*, 569. [CrossRef] [PubMed]
8. Bosquet, L.; Léger, L.; Legros, P. Blood lactate response to overtraining in male endurance athletes. *Eur. J. Appl. Physiol.* **2001**, *84*, 107–114. [CrossRef] [PubMed]
9. Kobayashi, Y.; Takeuchi, T.; Hosoi, T.; Yoshizaki, H.; Loeppky, J.A. Effect of a marathon run on serum lipoproteins, creatine kinase, and lactate dehydrogenase in recreational runners. *Res. Q. Exerc. Sport* **2005**, *76*, 450–455. [CrossRef]
10. Billat, L.V.; Koralsztein, J.P. Significance of the velocity at VO 2max and time to exhaustion at this velocity. *Sports Med.* **1996**, *22*, 90–108. [CrossRef] [PubMed]
11. Cannon, D.T.; White, A.C.; Andriano, M.F.; Kolkhorst, F.W.; Rossiter, H.B. Skeletal muscle fatigue precedes the slow component of oxygen uptake kinetics during exercise in humans. *J. Physiol.* **2011**, *589*, 727–739. [CrossRef] [PubMed]
12. Robson-Ansley, P.J.; Gleeson, M.; Ansley, L. Fatigue management in the preparation of Olympic athletes. *J. Sports Sci.* **2009**, *27*, 1409–1420. [CrossRef]
13. Lee, S.M.; Yoon, S.M.; Cho, H. Human activity recognition from accelerometer data using Convolutional Neural Network. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, Korea, 13–16 February 2017; pp. 131–134.
14. Nithya, N.; Nallavan, G. Role of Wearables in Sports based on Activity recognition and biometric parameters: A Survey. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 1700–1705.
15. Op De Beéck, T.; Meert, W.; Schütte, K.; Vanwanseele, B.; Davis, J. Fatigue prediction in outdoor runners via machine learning and sensor fusion. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 606–615.
16. Barshan, B.; Yüksek, M.C. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput. J.* **2014**, *57*, 1649–1667. [CrossRef]
17. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J. Fusion of smartphone motion sensors for physical activity recognition. *Sensors* **2014**, *14*, 10146–10176. [CrossRef] [PubMed]
18. Lockhart, J.W.; Weiss, G.M. Limitations with activity recognition methodology & data sets. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, Seattle, WA, USA, 13–17 September 2014; pp. 747–756.
19. Barshan, B.; Yurtman, A. Investigating inter-subject and inter-activity variations in activity recognition using wearable motion sensors. *Comput. J.* **2016**, *59*, 1345–1362. [CrossRef]
20. Kristiansen, M.; Madeleine, P.; Hansen, E.A.; Samani, A. Inter-subject variability of muscle synergies during bench press in power lifters and untrained individuals. *Scand. J. Med. Sci. Sports* **2015**, *25*, 89–97. [CrossRef]
21. Mourão-Miranda, J.; Hardoon, D.R.; Hahn, T.; Marquand, A.F.; Williams, S.C.; Shawe-Taylor, J.; Brammer, M. Patient classification as an outlier detection problem: An application of the one-class support vector machine. *Neuroimage* **2011**, *58*, 793–804. [CrossRef] [PubMed]
22. Kobsar, D.; Ferber, R. Wearable sensor data to track subject-specific movement patterns related to clinical outcomes using a machine learning approach. *Sensors* **2018**, *18*, 2828. [CrossRef] [PubMed]
23. Lubetzky-Vilnai, A.; Ciol, M.; McCoy, S.W. Statistical analysis of clinical prediction rules for rehabilitation interventions: Current state of the literature. *Arch. Phys. Med. Rehabil.* **2014**, *95*, 188–196. [CrossRef]
24. Fredriksson, T.; Mattos, D.I.; Bosch, J.; Olsson, H.H. Data labeling: An empirical investigation into industrial challenges and mitigation strategies. In Proceedings of the International Conference on Product-Focused Software Process Improvement, Turin, Italy, 25–27 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 202–216.
25. Nweke, H.F.; Teh, Y.W.; Al-Garadi, M.A.; Alo, U.R. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Syst. Appl.* **2018**, *105*, 233–261. [CrossRef]

26. Fallahzadeh, R.; Ghasemzadeh, H. Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data. In Proceedings of the 8th International Conference on Cyber-Physical Systems, Pittsburgh, PA, USA, 18–21 April 2017; pp. 293–302.

27. Sztyler, T.; Stuckenschmidt, H. Online personalization of cross-subjects based activity recognition models on wearable devices. In Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom), Kona, HI, USA, 13–17 March 2017; pp. 180–189.

28. Lane, N.D.; Xu, Y.; Lu, H.; Hu, S.; Choudhury, T.; Campbell, A.T.; Zhao, F. Enabling large-scale human activity inference on smartphones using community similarity networks (csn). In Proceedings of the 13th International Conference on Ubiquitous Computing, Beijing, China, 17–21 September 2011; pp. 355–364.

29. Chen, Y.; Wang, J.; Huang, M.; Yu, H. Cross-position activity recognition with stratified transfer learning. *Pervasive Mob. Comput.* **2019**, *57*, 1–13. [CrossRef]

30. Khan, M.A.A.H.; Roy, N.; Misra, A. Scaling human activity recognition via deep learning-based domain adaptation. In Proceedings of the 2018 IEEE international conference on pervasive computing and communications (PerCom), Athens, Greece, 19–23 March 2018; pp. 1–9.

31. Ferrari, A.; Micucci, D.; Mobilio, M.; Napoletano, P. On the personalization of classification models for human activity recognition. *IEEE Access* **2020**, *8*, 32066–32079. [CrossRef]

32. Palmius, N.; Saunders, K.E.; Carr, O.; Geddes, J.R.; Goodwin, G.M.; De Vos, M. Group-personalized regression models for predicting mental health scores from objective mobile phone data streams: Observational study. *J. Med Internet Res.* **2018**, *20*, e10194. [CrossRef] [PubMed]

33. Huang, Z.; Niu, Q.; You, I.; Pau, G. Acceleration Feature Extraction of Human Body Based on Wearable Devices. *Energies* **2021**, *14*, 924. [CrossRef]

34. Janidarmian, M.; Roshan Fekr, A.; Radecka, K.; Zilic, Z. A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors* **2017**, *17*, 529. [CrossRef]

35. Mannini, A.; Intille, S.S. Classifier personalization for activity recognition using wrist accelerometers. *IEEE J. Biomed. Health Informatics* **2018**, *23*, 1585–1594. [CrossRef]

36. Weiss, G.M.; Lockhart, J. The impact of personalization on smartphone-based activity recognition. In Proceedings of the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.

37. Hong, J.H.; Ramos, J.; Dey, A.K. Toward Personalized Activity Recognition Systems With a Semipopulation Approach. *IEEE Trans. Hum. -Mach. Syst.* **2016**, *46*, 101–112. [CrossRef]

38. Igual, R.; Medrano, C.; Plaza, I. A comparison of public datasets for acceleration-based fall detection. *Med. Eng. Phys.* **2015**, *37*, 870–878. [CrossRef] [PubMed]

39. Steffen, L.M.; Arnett, D.K.; Blackburn, H.; Shah, G.; Armstrong, C.; Luepker, R.V.; Jacobs, J.D. Population trends in leisure-time physical activity: Minnesota Heart Survey, 1980–2000. *Med. Sci. Sports Exerc.* **2006**, *38*, 1716–1723. [CrossRef] [PubMed]

40. Troiano, R.P.; Berrigan, D.; Dodd, K.W.; Masse, L.C.; Tilert, T.; McDowell, M. Physical activity in the United States measured by accelerometer. *Med. Sci. Sports Exerc.* **2008**, *40*, 181–188. [CrossRef]

41. Mair, S.D.; Seaber, A.V.; Glisson, R.R.; Garrett, W.E., Jr. The role of fatigue in susceptibility to acute muscle strain injury. *Am. J. Sports Med.* **1996**, *24*, 137–143. [CrossRef]

42. Garrett, W.E., Jr. Muscle strain injuries. *Am. J. Sports Med.* **1996**, *24*, S2–S8. [CrossRef]

43. Nesterenko, S.; Domire, Z.J.; Morrey, B.F.; Sanchez-Sotelo, J. Elbow strength and endurance in patients with a ruptured distal biceps tendon. *J. Shoulder Elb. Surg.* **2010**, *19*, 184 –189. [CrossRef]

44. Hopkins, W.G.; Marshall, S.W.; Quarrie, K.L.; Hume, P.A. Risk factors and risk statistics for sports injuries. *Clin. J. Sport Med.* **2007**, *17*, 208–210.

45. Elshafei, M.; Shihab, E. Towards Detecting Biceps Muscle Fatigue in Gym Activity Using Wearables. *Sensors* **2021**, *21*, 759. [CrossRef]

46. Apriantono, T.; Nunome, H.; Ikegami, Y.; Sano, S. The effect of muscle fatigue on instep kicking kinetics and kinematics in association football. *J. Sports Sci.* **2006**, *24*, 951–960. [CrossRef]

47. Sanchez-Medina, L.; González-Badillo, J.J. Velocity loss as an indicator of neuromuscular fatigue during resistance training. *Med. Sci. Sports Exerc.* **2011**, *43*, 1725–1734. [CrossRef]

48. Coelho, A.C.; Cannon, D.T.; Cao, R.; Porszasz, J.; Casaburi, R.; Knorst, M.M.; Rossiter, H.B. Instantaneous quantification of skeletal muscle activation, power production, and fatigue during cycle ergometry. *J. Appl. Physiol.* **2015**, *118*, 646–654. [CrossRef]

49. Borg, G.A. Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc.* **1982**, *14*, 377–381. [CrossRef]

50. Yoo, S.; Ackad, C.; Heywood, T.; Kay, J. Evaluating the actual and perceived exertion provided by virtual reality games. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 3050–3057.

51. Liao, F.; Zhang, X.; Cao, C.; Hung, I.Y.J.; Chen, Y.; Jan, Y.K. Effects of muscle fatigue and recovery on complexity of surface electromyography of Biceps Brachii. *Entropy* **2021**, *23*, 1036. [CrossRef]

52. Hwang, H.J.; Chung, W.H.; Song, J.H.; Lim, J.K.; Kim, H.S. Prediction of biceps muscle fatigue and force using electromyography signal analysis for repeated isokinetic dumbbell curl exercise. *J. Mech. Sci. Technol.* **2016**, *30*, 5329–5336. [CrossRef]

53. Bergquist, R.; Iversen, V.M.; Mork, P.J.; Fimland, M.S. Muscle activity in upper-body single-joint resistance exercises with elastic resistance bands vs. free weights. *J. Hum. Kinet.* **2018**, *61*, 5. [CrossRef]

54. Reis, V.M.; Garrido, N.D.; Vianna, J.; Sousa, A.C.; Alves, J.V.; Marques, M.C. Energy cost of isolated resistance exercises across low-to high-intensities. *PLoS ONE* **2017**, *12*, e0181311. [CrossRef]

55. Elshafei, M.; Costa, D.E.; Shihab, E. On the Impact of Biceps Muscle Fatigue in Human Activity Recognition. *Sensors* **2021**, *21*, 1070. [CrossRef]

56. Borg, G. *Borg's Perceived Exertion and Pain Scales.*; Human Kinetics: Champaign, IL, USA, 1998; pp. 1–97.

57. Adirim, T.A.; Cheng, T.L. Overview of injuries in the young athlete. *Sports Med.* **2003**, *33*, 75–81. [CrossRef]

58. Burt, C.W.; Overpeck, M.D. Emergency visits for sports-related injuries. *Ann. Emerg. Med.* **2001**, *37*, 301–308. [CrossRef]

59. Green, B.; Pizzari, T. Calf muscle strain injuries in sport: A systematic review of risk factors for injury. *Br. J. Sports Med.* **2017**, *51*, 1189–1194. [CrossRef]

60. Prentice, A.M.; Jebb, S.A. Beyond body mass index. *Obes. Rev.* **2001**, *2*, 141–147. [CrossRef]

61. Borga, M.; West, J.; Bell, J.D.; Harvey, N.C.; Romu, T.; Heymsfield, S.B.; Leinhard, O.D. Advanced body composition assessment: from body mass index to body composition profiling. *J. Investig. Med.* **2018**, *66*, 1–9. [CrossRef]

62. Burkhauser, R.V.; Cawley, J. Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *J. Health Econ.* **2008**, *27*, 519–529. [CrossRef]

63. Jones, B.H.; Hauret, K.G.; Dye, S.K.; Hauschild, V.D.; Rossi, S.P.; Richardson, M.D.; Friedl, K.E. Impact of physical fitness and body composition on injury risk among active young adults: A study of army trainees. *J. Sci. Med. Sport* **2017**, *20*, S17–S22. [CrossRef]

64. Janssen, I.; Katzmarzyk, P.T.; Ross, R. Body mass index, waist circumference, and health risk: Evidence in support of current National Institutes of Health guidelines. *Arch. Intern. Med.* **2002**, *162*, 2074–2079. [CrossRef]

65. Whittaker, R.L.; Sonne, M.W.; Potvin, J.R. Ratings of perceived fatigue predict fatigue induced declines in muscle strength during tasks with different distributions of effort and recovery. *J. Electromyogr. Kinesiol.* **2019**, *47*, 88–95. [CrossRef]

66. Alarfaj, M.; Qian, Y.; Liu, H. Detection of Human Body Movement Patterns Using IMU and Barometer. In Proceedings of the 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), Sharjah, United Arab Emirates, 16–18 March 2021; pp. 1–6.

67. Webber, M.; Rojas, R.F. Human Activity Recognition with Accelerometer and Gyroscope: A Data Fusion Approach. *IEEE Sens. J.* **2021**, *21*, 16979–16989. [CrossRef]

68. Li, Q.; Li, R.; Ji, K.; Dai, W. Kalman filter and its application. In Proceedings of the 2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS), Tianjin, China, 1–3 November 2015; pp. 74–77.

69. Wichit, N.; Choksuriwong, A. Multi-sensor data fusion model based Kalman filter using fuzzy logic for human activity detection. *Int. J. Inf. Electron. Eng.* **2015**, *5*, 450. [CrossRef]

70. Palumbo, F.; Gallicchio, C.; Pucci, R.; Micheli, A. Human activity recognition using multisensor data fusion based on reservoir computing. *J. Ambient. Intell. Smart Environ.* **2016**, *8*, 87–107. [CrossRef]

71. Wang, Y.; Cang, S.; Yu, H. A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition. *IEEE Sens. J.* **2018**, *18*, 6874–6888. [CrossRef]

72. Nweke, H.F.; Teh, Y.W.; Mujtaba, G.; Al-Garadi, M.A. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Inf. Fusion* **2019**, *46*, 147–170. [CrossRef]

73. Ferrari, A.; Micucci, D.; Mobilio, M.; Napoletano, P. Hand-crafted features vs residual networks for human activities recognition using accelerometer. In Proceedings of the 2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT), Ancona, Italy, 19–21 June 2019; pp. 153–156.

74. Vanrell, S.R.; Milone, D.H.; Rufiner, H.L. Assessment of homomorphic analysis for human activity recognition from acceleration signals. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1001–1010. [CrossRef]

75. Bianco, S.; Napoletano, P.; Schettini, R. Multimodal car driver stress recognition. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, Trento, Italy, 20–23 May 2019; pp. 302–307.

76. Ji, Q.; Lan, P.; Looney, C. A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2006**, *36*, 862–875.

77. Mallis, M.M.; Mejdal, S.; Nguyen, T.T.; Dinges, D.F. Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviat. Space Environ. Med.* **2004**, *75*, A4–A14.

78. Sant'Ana, M.; Li, G.; Zhang, H. A decentralized sensor fusion approach to human fatigue monitoring in maritime operations. In Proceedings of the 2019 IEEE 15th International Conference on Control and Automation (ICCA), Edinburgh, UK, 16–19 July 2019; pp. 1569–1574.

79. Aghamohammadi-Sereshki, A.; Bayazi, M.J.D.; Ghomsheh, F.T.; Amirabdollahian, F. Investigation of Fatigue Using Different EMG Features. In Proceedings of the 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), Toronto, ON, Canada, 24–28 June 2019; pp. 115–120.

80. Morgan, P.T.; Smeuninx, B.; Breen, L. Exploring the impact of obesity on skeletal muscle function in older age. *Front. Nutr.* **2020**, *7*, 286. [CrossRef]

81. Tomlinson, D.J.; Erskine, R.M.; Morse, C.I.; Pappachan, J.M.; Sanderson-Gillard, E.; Onambélé-Pearson, G.L. The combined effects of obesity and ageing on skeletal muscle function and tendon properties in vivo in men. *Endocrine* **2021**, *72*, 411–422. [CrossRef]

82. Malkauthekar, M. Analysis of Euclidean distance and Manhattan distance measure in Face recognition. In Proceedings of the Third International Conference on Computational Intelligence and Information Technology (CIIT 2013), Mumbai, India, 18–19 October 2013; pp. 503–507.

83. Aggarwal, C.C.; Hinneburg, A.; Keim, D.A. On the surprising behavior of distance metrics in high dimensional space. In Proceedings of the International Conference on Database Theory, London, UK, 4–6 January 2001; pp. 420–434.

84. Shirkhorshidi, A.S.; Aghabozorgi, S.; Wah, T.Y. A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS ONE* **2015**, *10*, e0144059. [CrossRef]

85. Shepard, R.N. Toward a universal law of generalization for psychological science. *Science* **1987**, *237*, 1317–1323. [CrossRef]

86. Tenenbaum, J.B.; Griffiths, T.L. Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* **2001**, *24*, 629–640. [CrossRef]

87. Walse, K.H.; Dharaskar, R.V.; Thakare, V.M. A study on the effect of adaptive boosting on performance of classifiers for human activity recognition. In Proceedings of the International Conference on Data Engineering and Communication Technology, Maharashtra, India, 15–16 December 2017; Springer: Singapore, 2017; pp. 419–429.

88. Vezhnevets, A.; Vezhnevets, V. Modest AdaBoost-teaching AdaBoost to generalize better. In Proceedings of the Graphicon, Novosibirsk, Russia, 20–24 June 2005; Volume 12; pp. 987–997.

89. Solomatine, D.P.; Shrestha, D.L. AdaBoost. RT: A boosting algorithm for regression problems. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 1163–1168.

90. Jeong, G.M.; Truong, P.H.; Choi, S.I. Classification of three types of walking activities regarding stairs using plantar pressure sensors. *IEEE Sens. J.* **2017**, *17*, 2638–2639. [CrossRef]

91. Soro, A.; Brunner, G.; Tanner, S.; Wattenhofer, R. Recognition and repetition counting for complex physical exercises with deep learning. *Sensors* **2019**, *19*, 714. [CrossRef]

92. Subasi, A.; Dammas, D.H.; Alghamdi, R.D.; Makawi, R.A.; Albiety, E.A.; Brahimi, T.; Sarirete, A. Sensor based human activity recognition using adaboost ensemble classifier. *Procedia Comput. Sci.* **2018**, *140*, 104–111. [CrossRef]

93. Fang, H.; He, L.; Si, H.; Liu, P.; Xie, X. Human activity recognition based on feature selection in smart home using back-propagation algorithm. *ISA Trans.* **2014**, *53*, 1629–1638. [CrossRef]

94. Zhu, Y.; Wang, C.; Zhang, J.; Xu, J. Human activity recognition based on similarity. In Proceedings of the 2014 IEEE 17th International Conference on Computational Science and Engineering, Chengdu, China, 19–21 December 2014; pp. 1382–1387.

95. Ghazal, M.; Haeyeh, Y.A.; Abed, A.; Ghazal, S. Embedded Fatigue Detection Using Convolutional Neural Networks with Mobile Integration. In Proceedings of the 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Barcelona, Spain, 6–8 August 2018; pp. 129–133.

96. Lin, C.Y.; Marculescu, R. Model personalization for human activity recognition. In Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Austin, TX, USA, 23–27 March 2020; pp. 1–7.

97. Jebelli, H.; Lee, S. Feasibility of Wearable Electromyography (EMG) to Assess Construction Workers' Muscle Fatigue. In Proceedings of the Advances in Informatics and Computing in Civil and Construction Engineering, Chicago, IL, USA, 1–3 October 2019; Mutis, I., Hartmann, T., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 181–187.

98. Wan, S.; Qi, L.; Xu, X.; Tong, C.; Gu, Z. Deep learning models for real-time human activity recognition with smartphones. *Mob. Netw. Appl.* **2020**, *25*, 743–755. [CrossRef]

99. Gillinov, S.; Etiwy, M.; Wang, R.; Blackburn, G.; Phelan, D.; Gillinov, A.M.; Houghtaling, P.; Javadikasgari, H.; Desai, M.Y. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med. Sci. Sports Exerc.* **2017**, *49*, 1697–1703. [CrossRef]

100. Gil, E.; Orini, M.; Bailon, R.; Vergara, J.M.; Mainardi, L.; Laguna, P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiol. Meas.* **2010**, *31*, 1271. [CrossRef]

101. Schäfer, A.; Vagedes, J. How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *Int. J. Cardiol.* **2013**, *166*, 15–29. [CrossRef]

102. Shcherbina, A.; Mattsson, C.M.; Waggott, D.; Salisbury, H.; Christle, J.W.; Hastie, T.; Wheeler, M.T.; Ashley, E.A. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J. Pers. Med.* **2017**, *7*, 3. [CrossRef]

103. Falter, M.; Budts, W.; Goetschalckx, K.; Cornelissen, V.; Buys, R. Accuracy of Apple Watch Measurements for heart rate and energy expenditure in patients with cardiovascular disease: Cross-Sectional Study. *JMIR mHealth uHealth* **2019**, *7*, e11889. [CrossRef]

104. Arney, B.E.; Glover, R.; Fusco, A.; Cortis, C.; de Koning, J.J.; van Erp, T.; Jaime, S.; Mikat, R.P.; Porcari, J.P.; Foster, C. Comparison of RPE (rating of perceived exertion) scales for session RPE. *Int. J. Sport. Physiol. Perform.* **2019**, *14*, 994–996. [CrossRef]

105. Sala, E.; Lopomo, N.F.; Tomasi, C.; Romagnoli, F.; Morotti, A.; Apostoli, P.; De Palma, G. Importance of Work-Related Psychosocial Factors in Exertion Perception Using the Borg Scale Among Workers Subjected to Heavy Physical Work. *Front. Public Health* **2021**, *9*, 461. [CrossRef]

106. Wang, S.; Tang, H.; Wang, B.; Mo, J. A Novel Approach to Detecting Muscle Fatigue Based on sEMG by Using Neural Architecture Search Framework. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**. [CrossRef]

107. Gil-Martín, M.; San-Segundo, R.; Fernandez-Martinez, F.; Ferreiros-López, J. Improving physical activity recognition using a new deep learning architecture and post-processing techniques. *Eng. Appl. Artif. Intell.* **2020**, *92*, 103679. [CrossRef]