

Article

Deep Learning-Based Optimal Smart Shoes Sensor Selection for Energy Expenditure and Heart Rate Estimation

Heesang Eom ^{1,†}, Jongryun Roh ^{2,†}, Yuli Sun Hariyani ^{1,3} , Suwhan Baek ¹, Sukho Lee ⁴, Sayup Kim ^{2,*} and Cheolsoo Park ^{1,*} 

- ¹ Department of Computer Engineering, Kwangwoon University, Seoul 01897, Korea; 9200heesang@gmail.com (H.E.); yulisun@telkomuniversity.ac.id (Y.S.H.); zhszhsj@gmail.com (S.B.)
² Digital Transformation R&D Department, Korea Institute of Industrial Technology (KITECH), 143 Hangeulro, Ansan 15588, Korea; ssaccn@kitech.re.kr
³ School of Applied Science, Telkom University, Bandung 40257, Indonesia
⁴ Department of Leisure Sports, College of Ecological Environment, Kyungpook National University, Sangju-si 37224, Korea; ehduq132@gmail.com
* Correspondence: sayub@kitech.re.kr (S.K.); parkcheolsoo@kw.ac.kr (C.P.)
† These authors contributed equally to this work.

Abstract: Wearable technologies are known to improve our quality of life. Among the various wearable devices, shoes are non-intrusive, lightweight, and can be used for outdoor activities. In this study, we estimated the energy consumption and heart rate in an environment (i.e., running on a treadmill) using smart shoes equipped with triaxial acceleration, triaxial gyroscope, and four-point pressure sensors. The proposed model uses the latest deep learning architecture which does not require any separate preprocessing. Moreover, it is possible to select the optimal sensor using a channel-wise attention mechanism to weigh the sensors depending on their contributions to the estimation of energy expenditure (EE) and heart rate (HR). The performance of the proposed model was evaluated using the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). Moreover, the RMSE was 1.05 ± 0.15 , MAE 0.83 ± 0.12 and R^2 0.922 ± 0.005 in EE estimation. On the other hand, and RMSE was 7.87 ± 1.12 , MAE 6.21 ± 0.86 , and R^2 0.897 ± 0.017 in HR estimation. In both estimations, the most effective sensor was the z axis of the accelerometer and gyroscope sensors. Through these results, it is demonstrated that the proposed model could contribute to the improvement of the performance of both EE and HR estimations by effectively selecting the optimal sensors during the active movements of participants.

Keywords: smart shoe; energy expenditure; heart rate; channel wise attention; DenseNet; accelerometer; gyroscope; pressure sensor; deep learning



Citation: Eom, H.; Roh, J.; Hariyani, Y.S.; Baek, S.; Lee, S.; Kim, S.; Park, C. Deep Learning-Based Optimal Smart Shoes Sensor Selection for Energy Expenditure and Heart Rate Estimation. *Sensors* **2021**, *21*, 7058. <https://doi.org/10.3390/s21217058>

Academic Editor: Ki H. Chon

Received: 24 September 2021

Accepted: 21 October 2021

Published: 25 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wearable technologies have been continuously developed to improve the quality of human life and facilitate mobility and connectivity among users due to the rapid development of the Internet of Things (IoT). Its global demand is increasing every year [1–3]. Recently, several wearable devices, including wrist bands, watches, glasses, and shoes, have started enabling the continuous monitoring of an individual's health, wellness, and fitness [4]. In particular, the coronavirus disease (COVID-19) pandemic highlighted the importance of remote healthcare delivery, resulting in further expansion of the wearable technology market [3,5]. This is because wearable devices could continuously collect and analyze the movement and physiological data of a user and provide appropriate feedback in function of users' exercise information and health status.

The shoe is a useful wearable device that is easy to use, unobtrusive, lightweight, and easily available when doing outdoor activities [6–9]. Previous studies on shoes include gait type classification [9–11], step count [8,12,13], and energy expenditure (EE) estimation [14].

Three types of sensors (i.e., pressure, accelerometer, and gyroscope sensors) were equipped in the shoes to realize these tasks. These relatively low-cost sensors could be mounted in an unconstrained and convenient manner and record the movement information of users to estimate their physical behaviors.

The EE estimation was associated with physical activity (PA) which could influence an individual's health conditions [15]. The PA level, which can be quantitatively assessed, is highly correlated with the risk of developing cardiovascular diseases, diabetes, and obesity [16,17]. In addition, there are only a few studies conducted on EE estimation using shoes compared to those on gait type classification and step counting. In addition, the accelerometer is one of the most commonly used sensors in shoes and other various devices for estimating EE [18–22].

In a previous study, a regression model was designed to estimate personal characteristics such as age, gender, height, weight, and BMI using accelerometer sensor data [18,20]. On the other hand, Vathsangam et al. used an accelerometer and a gyroscope sensor together to estimate EE, showing the improvement of the EE estimation by utilizing both sensor data [23]. In addition, a pressure sensor can also provide significant information to estimate EE. In a study conducted by Ngueleu et al., they predicted the number of steps taken by users using pressure sensors that were equipped to their shoes [13]. The results show that there was a high correlation between the number of steps and EE conducted by Nielson et al. [19]. Moreover, the pressure sensor could also be used along with the accelerometer sensor to improve the EE estimation. In [22], EE was estimated using barometric pressure and triaxial accelerometer sensors in various states such as sitting, lying, and walking. Additionally, Sazonova et al. estimated EE using the data from the triaxial accelerometer and five pressure sensors which were measured whilst the participants performed various activities such as sitting, standing, walking, and cycling [14].

The World Health Organization (WHO) reported that more than 30% of fatalities worldwide are caused by cardiovascular diseases (CVDs) [24]. The heart rate variability (HRV) is known as an important risk index for CVDs [25]. Accordingly, in recent years, various types of wearable devices have been developed (e.g., a watch-type device mounting electrocardiogram (ECG) or photoplethysmogram (PPG) sensors) to conveniently measure heart rate (HR). However, in an exercise environment, ECG is inconvenient to measure and PPG is affected by severe noise due to the movement. Instead of measuring the direct cardiac response, Lee et al. estimated HR from the activity information measured using an accelerometer and gyroscope sensors attached to the chest [26,27].

In recent years, advanced deep learning algorithms have been developed with the help of increasing computing power and a sufficient big dataset. There have been studies on the application of the deep learning approach to the wearable technology [28–30], where the algorithm performed well in regression and classification problems using physiological sensor data [21,31,32]. Staudenmayer et al. reported that an artificial neural network (ANN) model can predict the EE information using the accelerometer signals [21]. However, they extracted hand-crafted features from the signals and fed them into the ANN model, which are challenging to extract and suboptimal in distinguishing sophisticated patterns in the signal due to its fixed model-based approach. Zhu et al. successfully improved the accuracy of the EE estimation using convolutional neural network (CNN) by extracting subtle patterns from the accelerometer and heart rate signals [33].

In the studies [23,33], the multichannel data from the accelerometer and gyroscope sensors were simultaneously analyzed to estimate EE and HR, which could have been improved by considering the significance of each channel data. It is important to investigate which channel's data are the most significant when multivariate input data can be obtained from multichannel sensors to derive the target variable. In recent studies, a method to determine the weight for each input channel to a neural network was suggested using the channel-wise attention based on deep learning techniques [34–36].

This study investigated the novel approach in estimating EE and HR using wearable sensors. A smart shoes system was selected for the convenience of users rather than the

direct cardiac response measurement system, owing to its unobtrusive and natural manner of measuring the activities of users in their daily life. Conventionally, smart shoes are equipped with three types of sensors (i.e., pressure, accelerometer, and gyroscope) to produce multichannel data. Moreover, a deep neural network model was designed to infer EE and HR information from the multichannel data without using model-based hand-crafted feature extraction methods, and the attention mechanism provides appropriate weights to the input channels of the networks to improve the inference performance. Additionally, the weights decided by the attention algorithm provide the importance of three different sensors and their channels to the estimation of the physiological variations, EE, and HR. This could also enhance our understanding of the designed deep neural network structure, also known as explainable artificial intelligence [37].

The rest of this study is organized as follows. Section 2 discusses the design and data collection process of the experiment. Section 3 introduces the structure and the learning process of the proposed deep learning model. In addition, Section 4 discusses the results of HR and EE estimations using the proposed model and statistical analysis of the attention weights of sensors used as inputs. The results presented in Section 4 are discussed in Section 5 using the existing related studies. Finally, this study is concluded in Section 6.

2. Materials and Methods

2.1. System Overview

Figure 1 shows the overall system architecture for EE and HR estimation. The participant in the study wore a calorimeter (K4b2, Cosmed, Italy) and a chest strap (H10, Polar, Finland) for EE and HR measurements. Moreover, for the signal detection of walking and running, four film-type pressure sensors on each foot and a sensor (BMI160, Bosch Corp, Reutlingen, Germany) capable of the simultaneous measurement of 3-axis accelerometers and gyroscopes were mounted between the shoe's insole and outsole (Salted, Korea). Their locations are shown in Figure 2. In the figure, the locations of the pressure sensors are illustrated on the anatomical sketch. All sensor signals were simultaneously measured as the participant ran on the treadmill and predicted the EE and HR by using the deep learning model. The predictions were evaluated using the measurements from the calorimeter and chest strap.

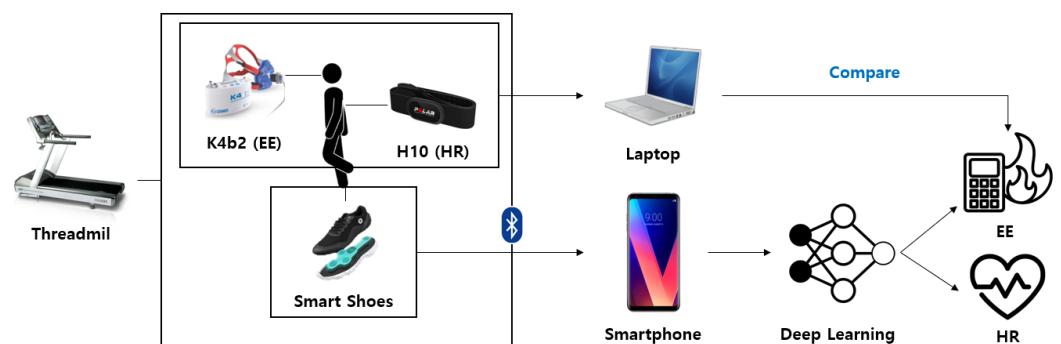


Figure 1. Overview of the system architecture for EE and HR estimation.

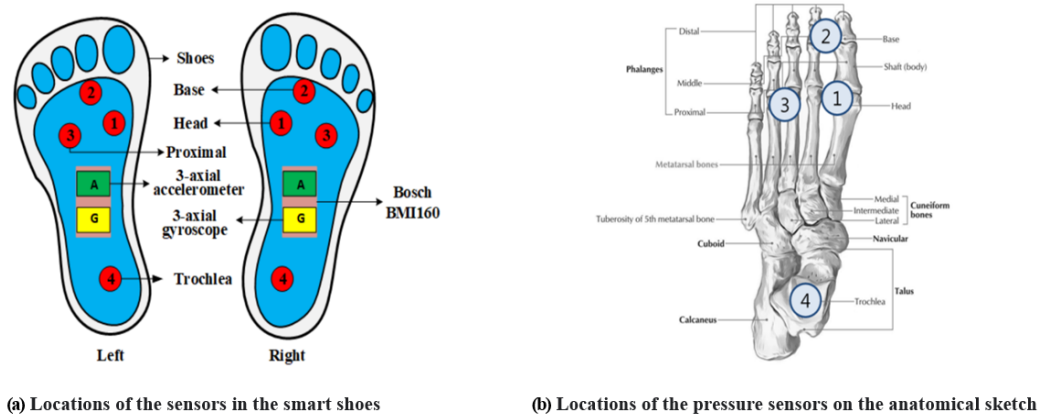


Figure 2. Locations of the sensors in the smart shoes: (a) a total of 12 sensors (6 sensors on the left and right shoe each) consisting of the pressure, accelerometer, and gyroscope sensors; (b) locations of the pressure sensors on the anatomical sketch: 1st metatarsal head (MH; sensor 1), toe (between the 1st and 2nd phalange; sensor 2), 4th metatarsal head (sensor 3), and heel (sensor 4).

2.2. Experiments

Ten healthy adult males (age: 22.5 ± 1.8 years old, height: 172.9 ± 3.5 cm, weight: 69.3 ± 4.9 kg, foot size: 264 ± 4.6 mm) without musculoskeletal and nervous system abnormalities were recruited for this study. Written informed consent was obtained from all participants. The study design and protocol was approved by the Institutional Review Board (IRB No. P01-201908-11-002).

The participants wore shoes equipped with pressure, accelerometer, and gyroscope sensors in their stable states before the experiment. In addition, as shown in Figure 3, they wore an HR strap and a calorimeter for measuring the HR and EE, respectively. Each participant ran on an electric treadmill at a speed varying from 3 to 10 kph, which increased by 1 kph per every 2 min (total 16 min ran) and they were instructed to run at a constant speed as much as possible. Each shoe data type of the participants (gyroscope, accelerometer, and pressure sensor data), HR, and EE were simultaneously recorded during the experiment. The shoes data were obtained using a smartphone app at a sampling rate of 33.3 Hz, while the HR and EE were acquired using the K4b2 software and recorded when the participant exhaled.



Figure 3. Example figures of the experimental equipment and process. During the experiment, participants wore a chest strap and a calorimeter to measure HR and EE, respectively. Each participant ran on a treadmill at a speed varying from 3 to 10 kph, which increased by 1 kph per every 2 min (total 16 min ran) and they were instructed to run at a constant speed as much as possible.

2.3. Data Preparation

Figure 4 shows the overall data preparation process for model training proposed in this study. It was difficult to determine the exact HR and EE that correspond to the data of the sensors attached on the shoes because the sampling periods of HR and EE recording (approximately 2–5 s) were not the same as those of the shoes' sensors (30 ms). Therefore, HR and EE data were resampled to match the sampling period of the data of the sensors attached on the shoe using a linear interpolation method, as shown in Figure 5. In addition, the obtained data were standardized for the efficient learning of the proposed deep learning model and reduced adverse effects of outliers [38]. The input sample used by the proposed deep learning model consisted of 20 channel data (four points' pressure on the left and right shoe each, triaxial accelerometer, and gyroscope) which were 10 s long, and the average values of HR and EE for 10 s were used as its label. The 10-s sample was overlapped to the next one by 1 s. The total number of samples was 9600. Moreover, Figure 6 shows the distributions of the HR and EE labels of 10 participants.

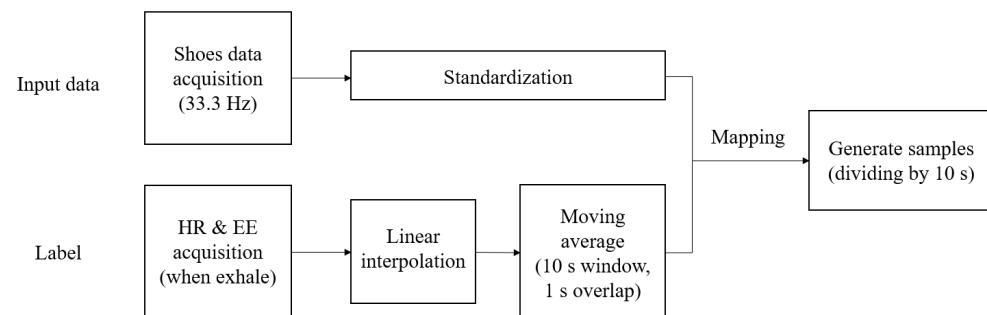


Figure 4. Flowchart of the data preprocessing when training the proposed deep learning model. The input shoes' data were recorded at a 33.3 Hz sampling rate and standardized to have a zero mean and unit variance. The label was created based on the HR and EE information, which were averaged on a 10 s long window with an overlap of 1 s.

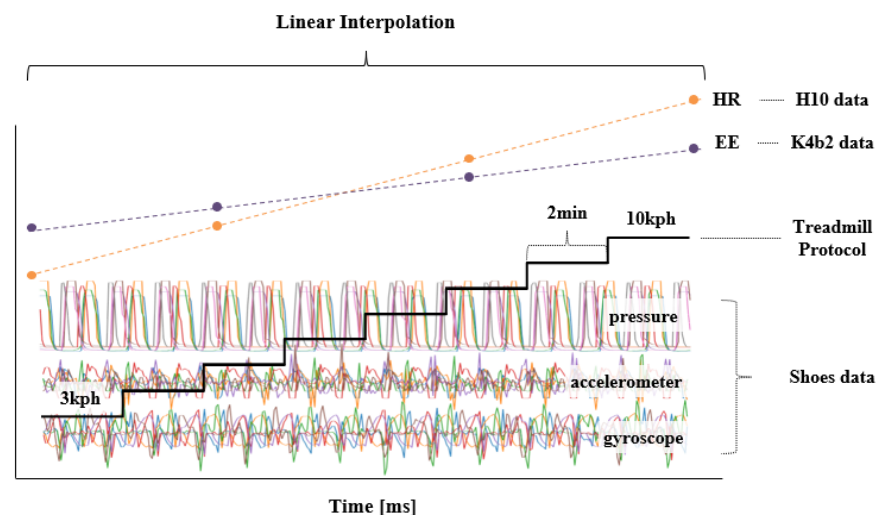


Figure 5. Application of a linear interpolation method due to the mismatch between the sampling rates of the HR/EE and data of the shoes' sensors. In the HR and EE graphs, the green dot represents HR and the gold dot represents the EE of the actual measurement, and the dashed line is the estimated value.

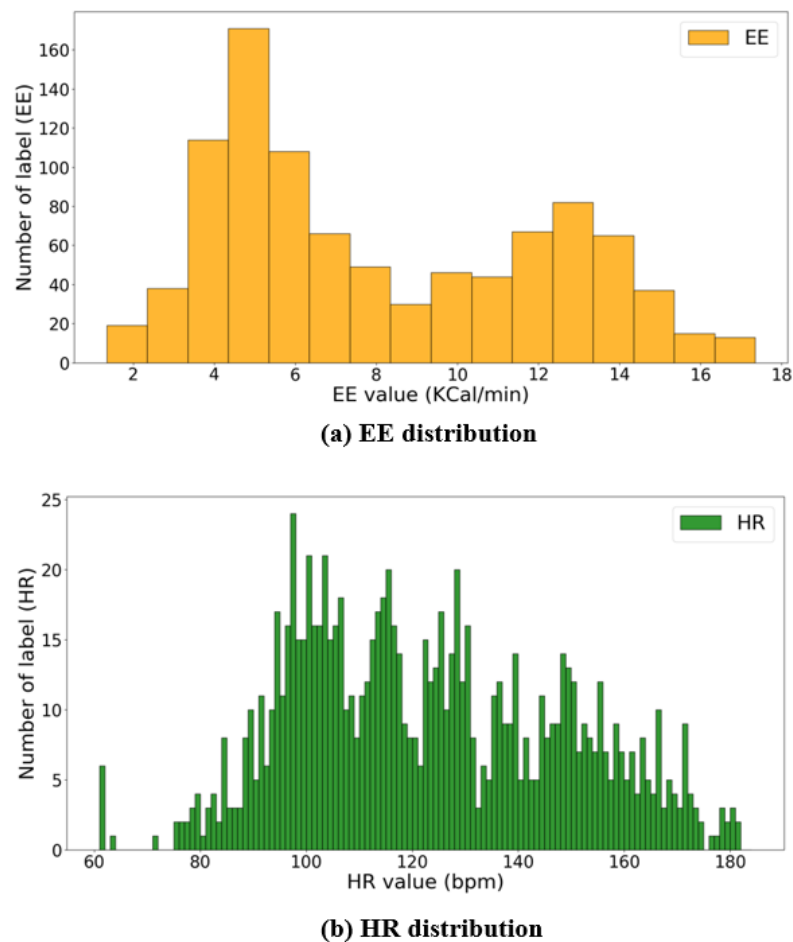


Figure 6. Distributions of HR and EE labels: (a,b) show the number of EE values per KCal/min and HR values per bpm, respectively.

3. Proposed Model

Figure 7 shows the overall structure of the model proposed in this study. The channel-wise attention layer, which is described in Section 3.1, provides weights to the significant channels of the sensors mounted on the shoes to accurately estimate HR and EE. The weighted signals by the attention layer pass using DenseNet [39], which is a CNN-based model known to be excellent in extracting key features from input data and generating spatial feature vectors that are discussed in Section 3.2. The bidirectional gated recurrent unit (GRU) [40] models the temporal relationship among the feature vectors, enabling an intuitive and efficient learning by observing the variations of input data over time (described in Section 3.3). Furthermore, the global average pooling (GAP) [41] layer compresses the information of the spatiotemporal features vectors and output values of HR and EE (described in Section 3.4). The advantages of the proposed model are as follows:

- The manual feature extraction process is not necessary since a fully automated end-to-end deep learning model was applied;
- The spatiotemporal characteristics of the multivariate time-series data that is complex to process could be effectively extracted using DenseNet and bidirectional GRU (Bi-GRU);
- The importance of each channel in estimating HR and EE could be quantified using the channel-wise attention method, and it can explain the optimal sensors for the task.

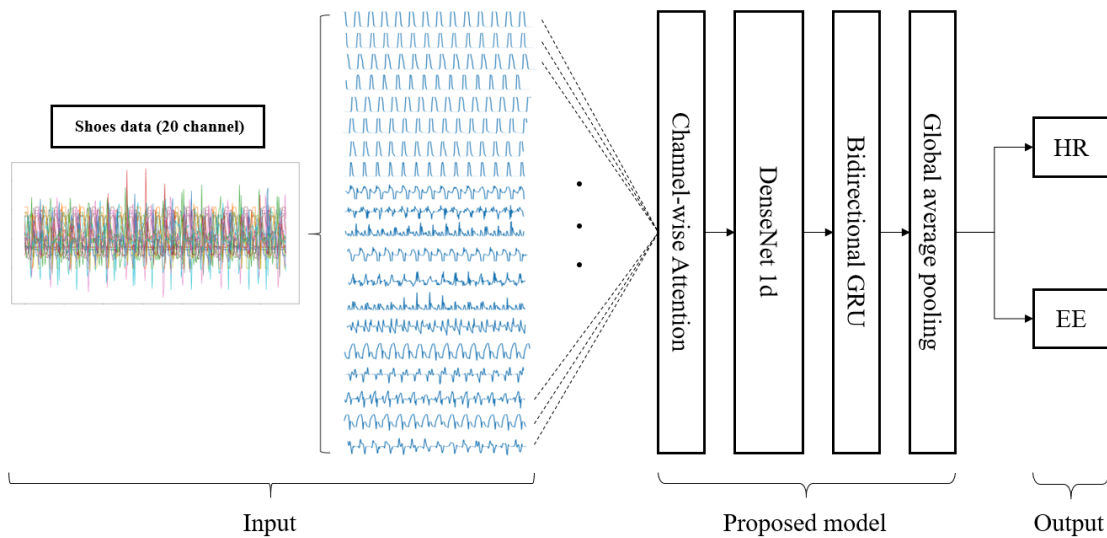


Figure 7. Structure of the proposed model. The shoe data from 20 channel sensors are fed into the input of the model and the channel-wise attention layer increases the intensity of the significant channels. The spatial features from the multi-channel data are extracted using DenseNet, and the temporal features are produced through Bi-GRU. Finally, HR and EE are estimated after the global average pooling (GAP) layer.

3.1. Channel-Wise Attention

It is difficult to extract the key features corresponding to HR and EE from the complex multivariate show data consisting of 20 channels. The conventional deep learning models train all input data with equal weights. This could deteriorate the learning efficiency of the model owing to the unnecessary and redundant information. However, the deep learning model could be efficiently trained by minimizing the unnecessary information in the input data and maximizing the significant information to the task. The attention mechanism is an optimized way of making this possible. In this study, we aimed to find and verify the optimal sensors for the estimation of HR and EE using the channel-wise attention expressed as follows:

$$O = \sigma(S_{in}W_{in} + b_{in}), \quad (1)$$

$$S_{att} = Att \otimes S_{in}, \quad (2)$$

$$Att = Average_t(O), \quad (3)$$

where O is calculated with the 20 channel signal $S_{in} = [s_1, s_2, \dots, s_i]$ ($s \in \mathbb{R}^t$, $S_{in} \in \mathbb{R}^{t \times i}$), a trainable weight matrix W_{in} ($W_{in} \in \mathbb{R}^{i \times i}$), a bias b_{in} ($b_{in} \in \mathbb{R}^t$), and a non-linear activation function $\sigma(\cdot)$. In addition, t is the time length of a sample and i the number of channels. A sigmoid function [42] was chosen in this study for the activation function. Att ($Att \in \mathbb{R}^t$) represents the attention weights, which is calculated by the average of O across the time axis using the $Average_t(\cdot)$ function. Finally, the signal S_{att} ($S_{att} \in \mathbb{R}^{t \times i}$) is derived by multiplying Att and S_{in} element-wise operation, which is expressed as \otimes .

3.2. DenseNet

DenseNet has yielded excellent performance in various image classification tasks [43–46]. Moreover, it avoids information dilution unlike other CNN-based models by concatenating the feature map output and input data in each convolutional layer. In addition, this method achieved higher performance with fewer parameters than that of the other models [39]. Therefore, DenseNet was used as a feature extractor in this study. The convolution layer was changed from two-dimensional (2D) to one-dimensional (1D), as shown in the Figure 8, since the shoe data are time-series data. In addition, the GAP layer was removed from its connection with the Bi-GRU layer in the last layer. The input to DenseNet S_{att} is produced from the channel-wise attention layer. The output is represented as follows:

$$F_{dense} = DenseNet(S_{att}) \quad (4)$$

The final output vector is $F_{dense} = [x_1, x_2, \dots, x_T]$ ($x \in \mathbb{R}^{c_1}, F_{dense} \in \mathbb{R}^{c_1 \times T}$), where T is the time length compressed by the pooling layer and c_1 is the number of output of the last convolution layer, because the DenseNet used in this study has no GAP in the last layer.

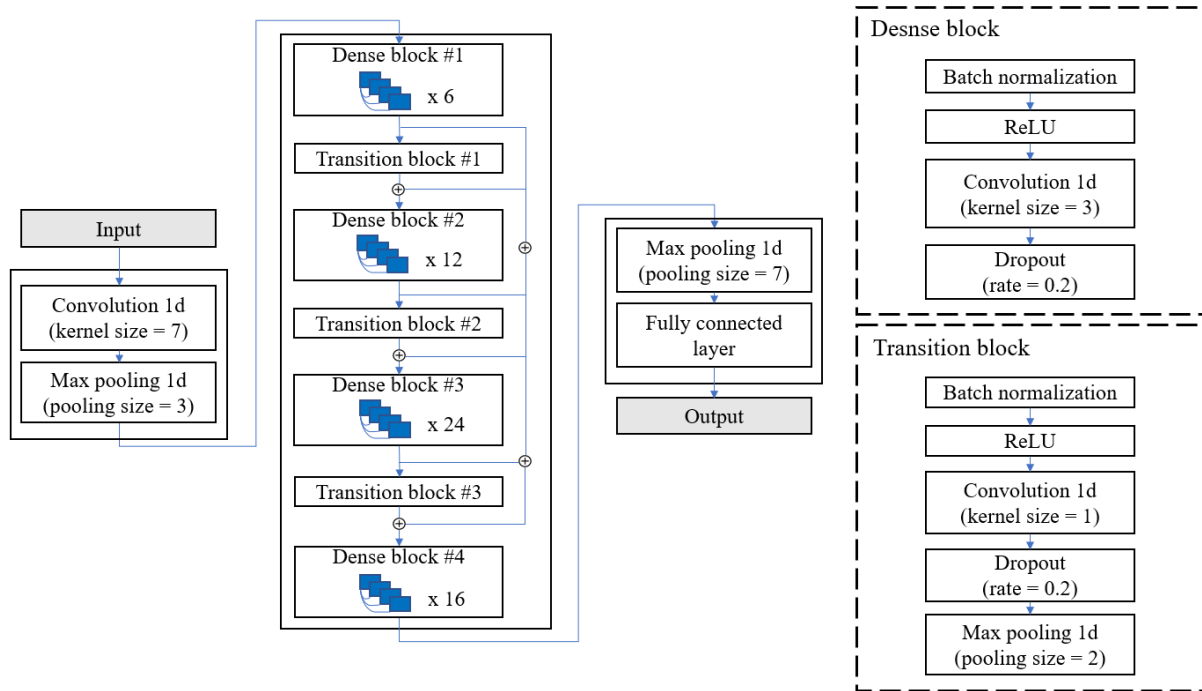


Figure 8. Internal structure of DenseNet. \oplus denotes the concatenation of feature vectors.

3.3. Bidirectional Gated Recurrent Unit

In the proposed model, the temporal features are extracted from the output of DenseNet, F_{dense} , using the Bi-GRU layer defined in Equation (5). GRU is one of RNN models with powerful modeling capabilities for long-term dependencies. On the other hand, long short-term memory (LSTM) [47] is another popular RNN model. Between the two, GRU has a more efficient structure with fewer parameters [40]:

$$F_{bigru} = BiGRU(F_{dense}) \quad (5)$$

The hidden vector of Bi-GRU, $F_{bigru} = [h_1, h_2, \dots, h_T]$ ($h \in \mathbb{R}^{c_2}, F_{bigru} \in \mathbb{R}^{c_2 \times T}$), was obtained from F_{dense} , where c_2 is the size of the hidden unit of the GRU, as shown in Figure 9. Moreover, the internal structure of the GRU cell is shown in Figure 10. The operation is elaborated as follows:

$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \quad (6)$$

$$z_t = \sigma(W_z * [h_{t-1}, x_t]) \quad (7)$$

$$\tilde{h}_t = \tanh(W_h * [r_t * h_{t-1}, x_t]) \quad (8)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (9)$$

In Equations (6)–(9), r_t and z_t are the update gate and the reset gate vectors for an arbitrary time point $t \in [1, T]$, respectively. The update gate determines how much information from the past and the present will be used to generate new information. The reset gate specifies which information to retain from the past information at the time $t - 1$. Moreover, \tilde{h}_t is a candidate state, which decides the amount of current information to be

learned using the result of the reset gate. W_z , W_r , and W_h are the trainable weight vectors of each gate. In addition, $\sigma(\cdot)$ and $\tanh(\cdot)$ are the sigmoid and hyperbolic tangential functions, respectively. Furthermore, $*$ denotes the element-wise multiplication.

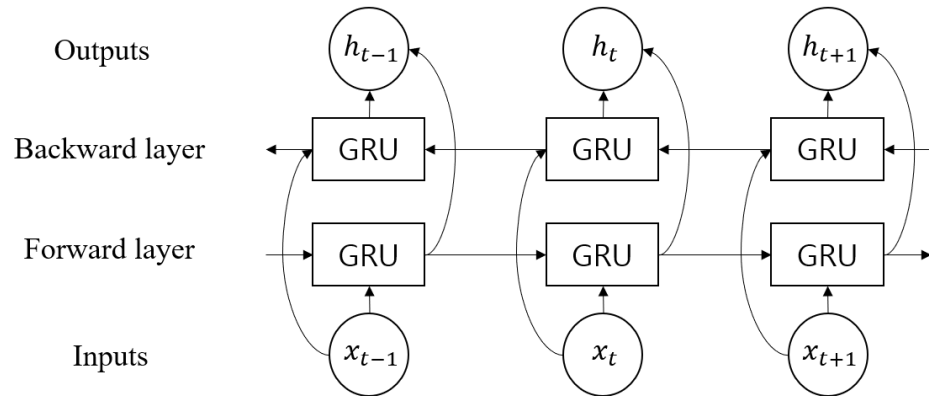


Figure 9. Structure of Bi-GRU.

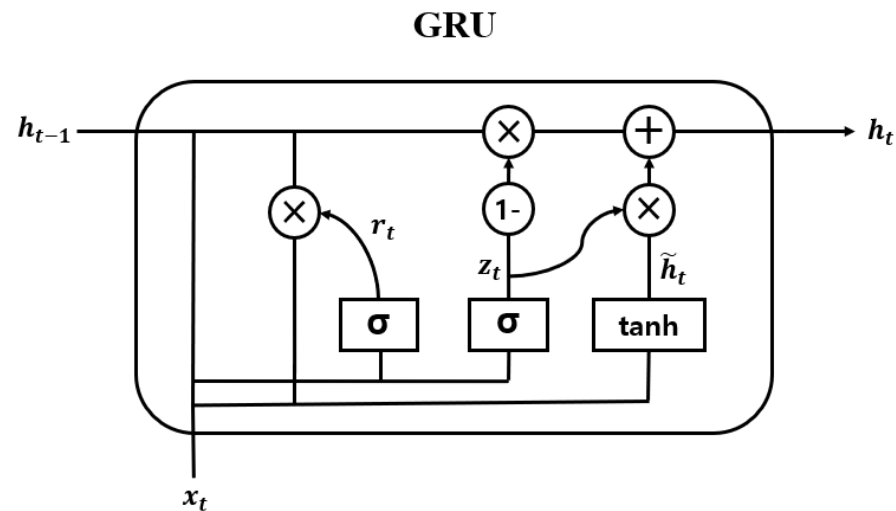


Figure 10. Internal structure of GRU.

Bi-GRU could simultaneously utilize both the past and future information, creating more useful features than unidirectional GRU. This is implemented as a forward and backward layer, as shown in Figure 9. The final output h_t of Bi-GRU is determined by the concatenation of the two vectors when the forward and backward hidden vectors are represented as \vec{h}_t and \overleftarrow{h}_t , respectively:

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (10)$$

3.4. Global Average Pooling

In the proposed model, the GAP layer was designed in the last layer instead of the fully connected (FC) layer, which tends to overfit on the training data. This could degrade the generalization performance of the networks. On the other hand, no additional parameters were required since the GAP layer only calculates the average across the final output vectors of the network, reducing the overall network size and preventing overfitting. The final predicted target variables (i.e., HR and EE) using GAP are calculated as follows:

$$Target = \left(\frac{1}{T} \sum_{t=1}^T F_{bigru} \right) W_{out} + b_{out}. \quad (11)$$

3.5. Model Training Environment

The proposed model uses leave-one-subject-out (LOSO) cross-validation to evaluate the robustness and generalizability in an inter-subject analysis. The data of 9 subjects out of 10 subjects were used as the training set and the data of the remaining 1 subject were used as the testing set, which was repeated for all subjects. The mean and standard deviation of performance for each subject were calculated and described in Section 4. The Adam [48] optimization (learning rate = 10^{-3}) was used to train the model, and the batch size was empirically set to 16. The initial weights of the networks were set at random and the loss function was designed based on the mean squared error (MSE). An early stopping method was applied to find the optimal model when there is no significant improvement in the validation loss of 20 epochs in a total of 150 training epochs. Furthermore, 4.2 GHz Intel Core i7 processor (Intel, Santa Clara, CA, USA) and NVIDIA GeForce RTX 2080Ti (NVIDIA corporation, Santa Clara, CA, USA) (with 11 GB VRAM), which are the computing environment for network training, were used. The model was implemented in Keras deep learning framework with TensorFlow backend.

4. Results

The results of the proposed model were evaluated in the following three aspects:

- Performance evaluation of the HR and EE estimation models;
- Performance analysis with and without the attention mechanism;
- Analysis of the channel significance using the attention weight;

The performance of the model was evaluated using several indicators. The root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) between the predicted and ground truths were calculated. Additionally, a Bland–Altman plot [49] was also presented. The formula of the evaluation indices are as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}, \quad (14)$$

In Equations (12)–(14), N is the total number of test samples, y_i is the ground truth, \hat{y}_i is the predicted value, and \bar{y}_i is the average value of y_i .

4.1. Energy Expenditure Estimation

4.1.1. Proposed Model Performance

Table 1 shows the EE estimation performance using the proposed model. The pressure, accelerometer, and gyroscope sensor data were all used as input data. The RMSE between the predicted and ground truths was 1.05 ± 0.13 , MAE was 0.83 ± 0.12 , and R^2 was 0.922 ± 0.005 . Figure 11 illustrates the predicted and ground truths across time for the best- and worst-case scenarios using the proposed model.

Table 1. EE (KCal/min) estimation performance.

Input	RMSE	MAE	R^2
Acc + Gyro + Pr	1.05 ± 0.13	0.83 ± 0.12	0.922 ± 0.005

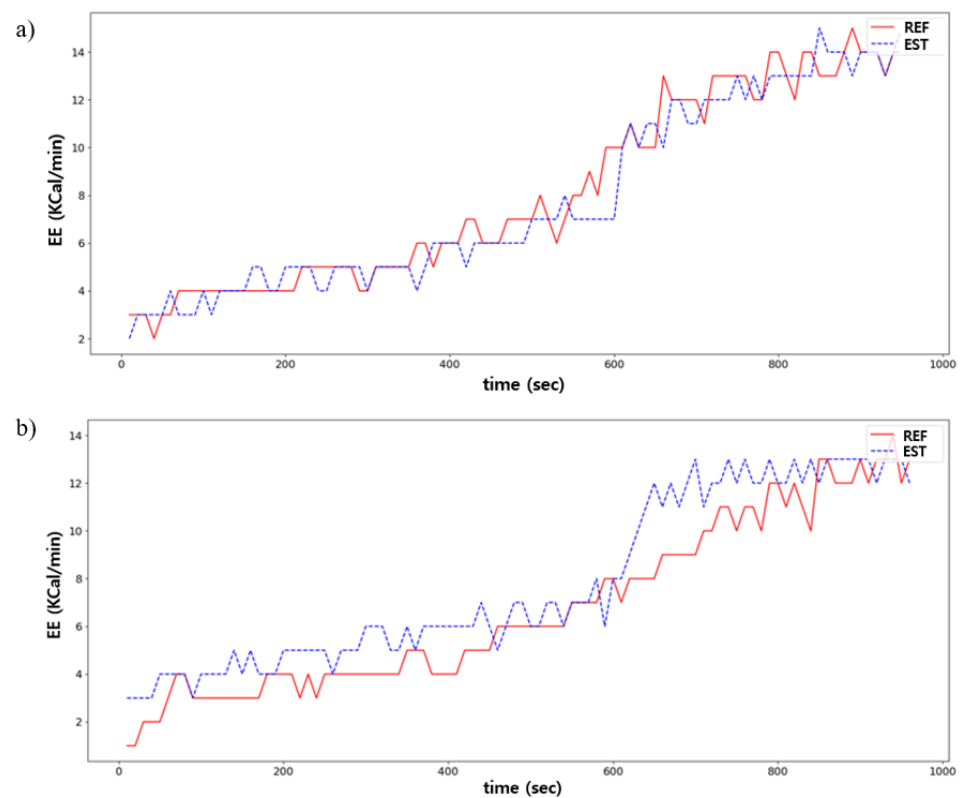


Figure 11. Comparison between the predicted (EST) and ground truths (REF) in EE estimation: (a) is the best case; (b) is the worst case.

4.1.2. Channel-Wise Attention Effectiveness

Analyzing what kind of sensors are helpful in estimating HR or EE using the channel-wise attention mechanism is the main objective of this study. This process could not be significant if the channel-wise attention degrades the performance of the model. The averaged results among the 10 participants are shown in Table 2 and Figure 12.

Table 2. Mean and standard deviation of RMSE, MAE, and R^2 values obtained using the proposed models with and without the attention mechanism in the EE estimation.

Input	RMSE	MAE	R^2
with attention (proposed)	1.05 ± 0.13	0.83 ± 0.12	0.922 ± 0.005
without attention	1.17 ± 0.24	0.95 ± 0.2	0.923 ± 0.12

The proposed model using the channel-wise attention in EE estimation achieved higher performance in RMSE and MAE compared to that without the channel-wise attention. In addition, a significant improvement was confirmed using the one-tailed paired-sample t -tests ($p < 0.05$). Although the R^2 of both models were quite close, the performance of the attention model is more stable with a standard deviation that is less than 0.005.

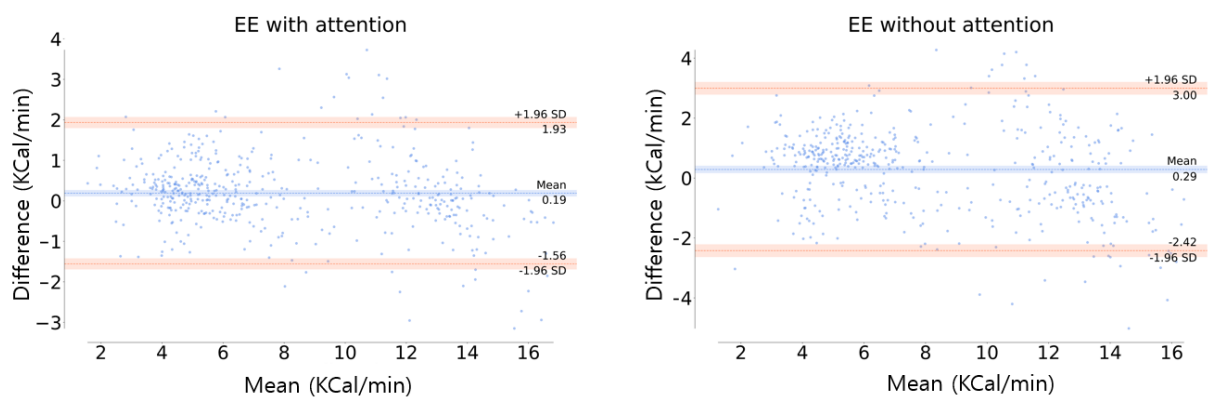


Figure 12. Bland-Altman plot of EE estimation. The orange line is the limit of agreement (LOA) and the center blue line is the mean of difference error between the actual and estimation.

In Figure 12, the orange line is the limit of agreement (LOA), which means that the difference between the actual and estimated values is within the range of [lower LOA, upper LOA]. At the 95% confidence interval ($\pm 1.96SD$), in this study, the LOA was $[-1.56, 1.93]$ and $[-2.42, 3]$ in the model with and without the channel-wise attention, respectively. The distribution of the difference values was more concentrated around zero in the model with the attention than the model without attention, indicating the superiority of the channel-wise attention model. In addition, the mean differences between the actual and the estimation (blue line in Figure 12) were 0.19 and 0.29 for the models with and without the attention, respectively. This indicates the high accuracy of the attention model, which means that this model has little bias compared to the model without attention.

4.1.3. Optimal Sensor Analysis

The additional analysis was performed to determine the optimal sensors for the EE estimation based on the results presented above. First, one-way ANOVA was performed to investigate whether there is a significant difference among the average attention weights for each sensor calculated from the channel-wise attention. The results are shown in Table 3, where SS denotes the sum of squares, df denotes the degree of freedom, MS denotes the mean square, and F denotes the F-statistic. As a result of the ANOVA analysis, there was a statistically significant difference in the average attention weights of the sensors ($p < 0.001$). Therefore, we also conducted a post hoc Tukey HSD test and the results are shown in Table 4. In the post hoc analysis, the symbols *P*, *A*, and *G* represent pressure, accelerometer, and gyroscope, respectively. The first letter in the subscript denotes the left (*L*) or right (*R*) side of the shoe, and the second letter is the detailed attachment position of the pressure sensors (see Figure 2) or the *x*, *y*, and *z* axis of the accelerometer and gyroscope. Each numerical value is an attention weight, and each column corresponds to a homogeneous subset with no statistically significant difference. For example, in column 1 of Table 4, there is no statistical difference in attention weights from P_{L3} to P_{L1} sensors ($p > 0.05$), and the *p*-value of the corresponding subset is indicated at the bottom of the table. In Table 4, seven sensors are included in each subset from columns 1 to 6, but the number of sensors included in the subset sharply decrease in columns 7 to 10. This means that the sensors from columns 7 to 10 contributed significantly more to the EE estimation than those from columns 1 to 6. As a result, the subsets of sensors that are important for the EE estimation were $\{A_{LZ}, G_{LZ}, G_{LY}\}$, $\{G_{LY}, A_{RZ}\}$, and $\{A_{RZ}, G_{RZ}\}$ in the order of high attention weight. The accelerometer and gyroscope mostly show their higher contribution to the EE estimation than the pressure sensors, and particularly their attention weights in the *z* axis are higher than those in the other axes G_{LY} .

Table 3. ANOVA analysis of the channel-wise attention weights in the EE estimation. SS is the sum of squares, df is the degree of freedom, MS is the mean square, and F is the F-statistic.

	SS	df	MS	F	p-Value
between groups	1.216	19	0.064	55.107	0.000
within groups	22.434	19,320	0.001		
total	23.649	19,339			

Table 4. Post-hoc Tukey HSD test result for the averaged attention weight for each sensor in EE estimation. Each column 1–10 represents a homogeneous subset for a significance level of 0.05. The sensor types are pressure (P), accelerometer (A), and gyroscope (G). The first subscript for each sensor type denotes the left (L) and right (R) sides of the shoe. The second subscript is the detailed attachment position of the pressure sensor (see Figure 2) or the x, y, and z axis directions of the accelerometer and gyroscope.

Sensor Type	1	2	3	4	5	6	7	8	9	10
P_{L3}	0.4908									
P_{R2}	0.4910	0.4910								
P_{R4}	0.4910	0.4910								
P_{L4}	0.4918	0.4918								
G_{RY}	0.4926	0.4926	0.4926							
P_{R3}	0.4935	0.4935	0.4935							
P_{L1}	0.4944	0.4944	0.4944	0.4944						
A_{LY}		0.4964	0.4964	0.4964	0.4964					
P_{L2}			0.4977	0.4977	0.4977	0.4977				
A_{RX}			0.4979	0.4979	0.4979	0.4979				
A_{RY}			0.4980	0.4980	0.4980	0.4980				
P_{R1}				0.4991	0.4991	0.4991				
A_{LX}				0.4998	0.4998	0.4998				
G_{RX}					0.4999	0.4999				
G_{LX}						0.5031	0.5031			
G_{RZ}							0.5070	0.5070		
A_{RZ}								0.5091	0.5091	
G_{LY}									0.5137	0.5137
G_{LZ}										0.5148
A_{LZ}										0.5155
p-value	0.742	0.069	0.057	0.060	0.750	0.072	0.546	0.999	0.240	1.000

4.2. Heart Rate Estimation

4.2.1. Proposed Model Performance

There are few previous studies conducted about the HR estimation using the pressure, accelerometer, or gyroscope sensors compared with those about the EE estimation because it is relatively easy to obtain an accurate heart rate using various off-the-shelf wearable devices equipped with physiological sensors such as electrocardiogram (ECG) and photoplethysmogram (PPG). However, users might be uncomfortable wearing an additional wrist or chest band to measure ECG or PPG. On the other hand, shoes could be a natural and unobtrusive wearable device to measure users' physiological information. This study tried to extract the HR information from the pressure, accelerometer, and gyroscope sensors that were mounted on shoes due to the limitation of ECG and PPG measurements with high SNR by selecting the optimal sensors for the estimation. The performance of the heart rate estimation using the proposed model, which selects the optimal sensors with the help of the channel-wise attention mechanism, is shown in Table 5. Additionally, Figure 13 is the graph showing the actual and predicted values for the best and worst cases of the proposed model.

Table 5. HR (bpm) estimation performance.

Input	RMSE	MAE	R ²
Acc + Gyro + Pr	7.81 ± 1.12	6.12 ± 0.86	0.897 ± 0.017

Accurately measuring the heart rate using physical sensors attached to smart shoes is challenging. Since the purpose of this study is to make it possible to easily measure heart rate in daily life, we compared it with the heart rate estimation accuracy of PPG-based wearable devices that are commercially available. Table 6 lists the accuracy of consumer wearable devices in heart rate estimation conducted by Nelson et al. [50]. The two devices that were compared were Apple Watch 3 (2017 version, Apple Inc, Cupertino, CA, USA, v. 4.2.3) and Fitbit Charge 2 (2017 version, Fitbit Inc, CA, USA, v. 22.55.2). In addition, MAE, Bland–Altman analysis, and mean absolute percent error (MAPE) were calculated as performance evaluation metrics. In particular, MAPE was calculated as follows:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (15)$$

In the previous study conducted by Nelson et al., the performance of each device under various conditions was evaluated. However, in Table 6, only the results in walking and running environments similar to this study were compared. The performance of the proposed model was 5.40 of MAPE, which is good compared with the results of Fitbit Charge 2 (9.21 and 9.88) and slightly worse than that of Apple Watch 3 (4.61 and 3.01).

Table 6. Comparison of the HR estimation performance of commercial wearable devices and the proposed model. The performance of commercial wearable devices was cited from the study results of Nelson et al. [50].

Device	Condition	Device Error		Bland–Altman Analysis		
		MAE	MAPE	ME	Lower LOA	Upper LOA
Fitbit Charge 2	walking	9.55	9.21	−6.85	−28.51	14.81
	running	14.73	9.88	−14.73	−29.77	0.31
Apple Watch 3	walking	4.77	4.64	0.11	−14.18	14.41
	running	4.05	3.01	1.77	−9.78	13.33
proposed model	walking + running	6.12	5.40	0.39	−15.12	15.90

4.2.2. Channel-Wise Attention Effectiveness

The HR estimation with and without attention were also compared similar to EE estimation to verify the improvement of the performance using the channel-wise attention. The results are shown in Table 7. In HR estimation, the proposed model using the channel-wise attention achieved higher performance for all evaluation indicators including RMSE, MAE, and R², which are all statistically significant ($p < 0.05$). This indicates that the channel-wise attention contributes to the selection of the optimal sensor for estimating the correct HR.

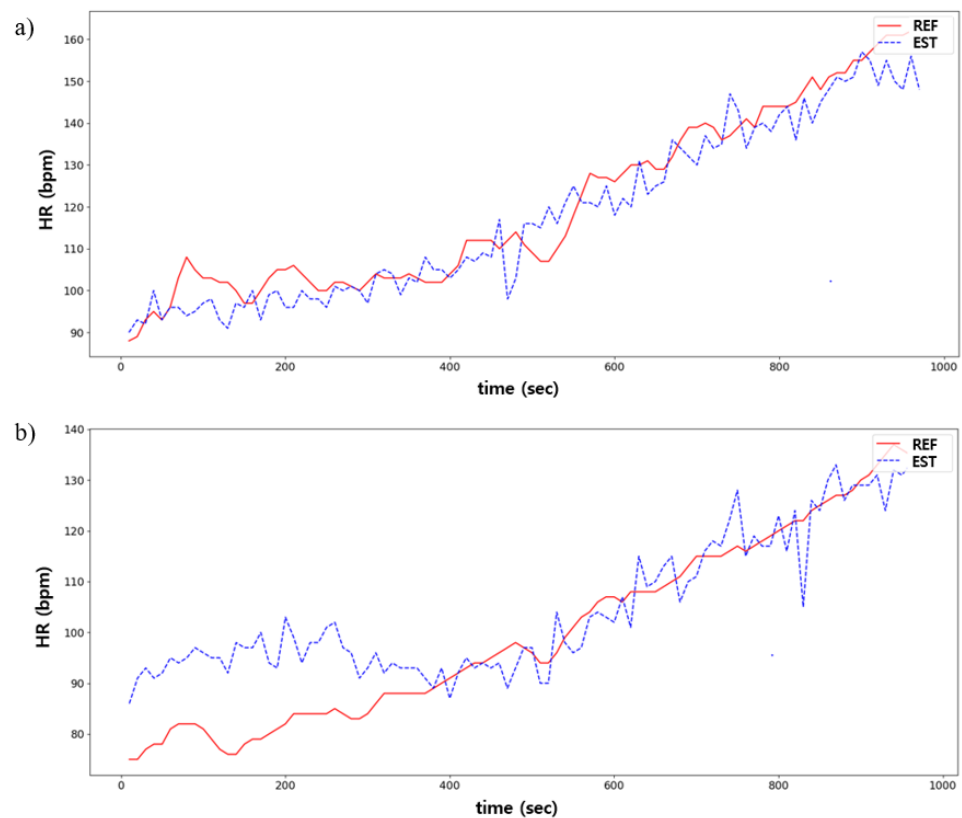


Figure 13. Comparison between predicted (EST) and ground truths (REF) in HR estimation: (a) is the best case; and (b) is the worst case.

Table 7. Mean and standard deviation of RMSE, MAE, and R^2 values of models with (proposed model) and without the attention for HR estimation.

Input	RMSE	MAE	R^2
with attention	7.81 ± 1.12	6.12 ± 0.86	0.897 ± 0.017
without attention	9.19 ± 3.16	7.72 ± 3.67	0.878 ± 0.037

Figure 14 shows the Bland–Altman plot of HR estimation, where the LOA was in the range of $[-15.12, 15.90]$ and $[-21.46, 16.79]$ in the model with and without the attention, respectively, at the 95% confidence interval ($\pm 1.96SD$). This indicates that the model with attention has less bias and higher stability than the model without attention, which is similar to the results of EE estimation.

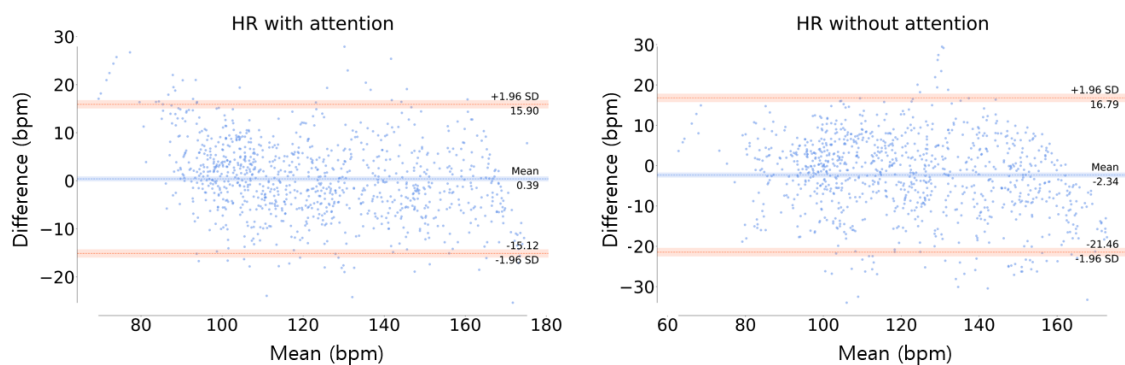


Figure 14. Bland–Altman plot of HR estimation. The orange line represents the limit of agreement (LOA) and the blue center line is the mean of the difference error between the ground truth and the estimation.

4.2.3. Optimal Sensor Analysis

One-way ANOVA was performed for HR estimation in the same way as EE estimation, for which the results are shown in Table 8. As a result of ANOVA analysis, there was a statistically significant difference in the averaged attention weight between the sensors ($p < 0.001$). In addition, a post hoc Tukey HSD test was conducted, and the results are shown in Table 9. In the post hoc analysis, the homogeneous subsets that contributed to the HR estimation were shown in the following order: $\{A_{LZ}\}$, $\{A_{RZ}, G_{LY}\}$, $\{G_{LY}, G_{LZ}, G_{RZ}\}$. Same as the results of EE estimation, the accelerometer and gyroscope mostly showed a higher contribution than the pressure sensor and z axis direction sensors made a greater contribution than the other directions. In particular, the average attention weight of A_{LZ} was significantly different from those of the other sensors, followed by A_{RZ} .

Table 8. ANOVA analysis of the channel-wise attention weights in the HR estimation. SS denotes the sum of squares, df denotes the degree of freedom, MS denotes the mean square, and F denotes the F-statistic.

	SS	df	MS	F	p-Value
between groups	2.145	19	0.113	90.706	0.000
within groups	24.049	19,320	0.001		
total	26.194	19,339			

Table 9. Post-hoc Tukey HSD test result for the average attention weight for each sensor in HR estimation. Each column 1–8 represents a homogeneous subset for a significance level of 0.05.

Sensor Type	1	2	3	4	5	6	7	8
P_{R3}	0.4864							
P_{L3}	0.4871							
P_{R1}	0.4888							
P_{R4}	0.4893							
P_{R2}	0.4900	0.4900						
P_{L4}	0.4901	0.4901						
G_{RY}	0.4911	0.4911	0.4911					
A_{LX}		0.4952	0.4952					
A_{RY}		0.4954	0.4954					
P_{L1}		0.4956	0.4956					
A_{RX}			0.4961					
P_{L2}			0.4963					
G_{RX}				0.5032				
A_{LY}				0.5064	0.5064			
G_{LX}				0.5067	0.5067			
G_{RZ}				0.5084	0.5084	0.5084		
G_{LZ}					0.5099	0.5099		
G_{LY}						0.5141	0.5141	
A_{RZ}							0.5167	
A_{LZ}								0.5229
p-value	0.288	0.061	0.122	0.130	0.794	0.056	0.986	1.000

5. Discussion

In this study, it was shown that the proposed model could estimate the EE and HR using physical sensors such as accelerometer, gyroscope, and pressure sensors that can be equipped in smart shoes. In particular, the accuracy was improved with adaptively assigning weights to the sensors through the channel-wise attention, which is the core of the model to select the optimal sensors, making important contributions to the EE and HR estimations.

The proposed model shows that the z axis sensors in the accelerometer and gyroscope have higher contributions to the EE estimation than the others, as shown in Table 3 and Table 8. Among the previous EE estimation studies, Vathsangam et al. [23] calculated the EE in the treadmill while walking using an accelerometer sensor and a gyroscope sensor. They claimed that the x axis sensor in the accelerometer (y axis in this study) was aligned with the movement direction of the foot, indicating that its contribution to the EE estimation could be high. On the other hand, Javed et al. [51] found that the y and z axis features of the accelerometer were important to recognize walking and jogging activities. In another related study, Smith et al. [52] calculated the ratio of the triaxial to uniaxial (vertical) number in the accelerometer for various activities using an accelerometer sensor on the wrist. The results show that activities such as running are greatly affected by vertical movement. Moreover, we found that the average attention weight of the z axis was high corresponding to the running activity, which is largely affected by vertical activity. The findings of the significance of the z axis monitoring the vertical movement are consistent with the results of Javed et al. [51] and Smith et al. [52] since our study was conducted on a treadmill under similar conditions to the jogging activity.

In the HR estimation, the contributions of the z axis sensors in the accelerometer and gyroscope were high, which is similar to the results of EE estimation. In various previous EE estimation studies, the EE was directly calculated using the HR level [53]. However, in this study, the EE estimation was carried out separately from the HR estimation. As a result, large attention weights in the z axis in the proposed model seem to be significant considering the high correlation between HR and EE.

As an additional analysis, we performed ANOVA and post hoc analysis to verify whether there is a significant difference in attention weights among the x, y, and z axis sensors in the accelerometer and gyroscope. Figure 15 shows the average attention weight for each axis to predict the EE and HR levels. As a result, there was a significant difference between the x and z axes and between the y and z axes ($p < 0.001$), although there was no statistical difference between the x and y axes.

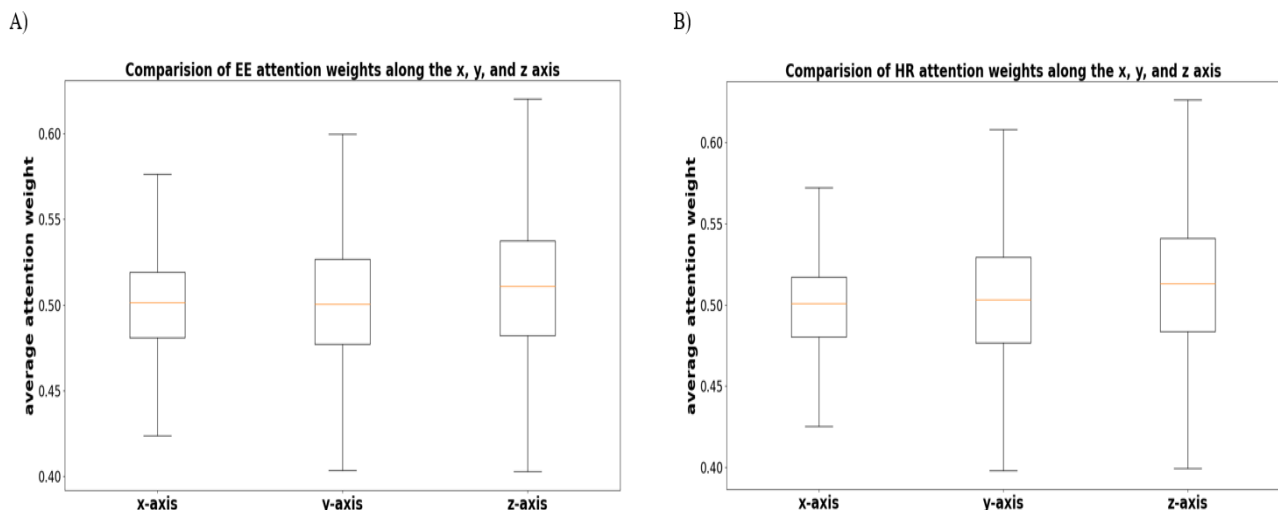


Figure 15. Comparison of the average attention weights for each of the x, y, and z axes. (A,B) illustrate the result of EE and HR, respectively.

6. Conclusions

In this study, the efficient HR and EE estimation models from multivariate raw signals including pressure, accelerometer, and gyroscope sensor data were designed using a deep learning architecture in an end-to-end manner. In addition, significant channels of the sensors were investigated using the channel-wise attention mechanism to estimate HR and EE, which found that the effects of the z axis sensors of the accelerometer and the gyroscope were significant in walking and running conditions. This is consistent with

the previous study demonstrating that a general running activity is greatly affected by a vertical movement in the z axis direction [51,52]. This study also demonstrated the possibility of estimating HR and EE using the sensors mounted on shoes and suggests an effective and cost-efficient design of a wearable shoe-based device with selecting the optimal sensors. Furthermore, using the channel-wise attention, HR and EE were effectively estimated even when the individual left and right foot movements were not constant during exercise. A limitation of this study is the small size of the training dataset and the individual characteristics of the participants with small deviations. Whilst the predictions might be a little unstable for datasets obtained under various conditions, the proposed model is trained and validated through the inter-subject analysis using LOSO, which could guarantee the generalizability of the proposed model if being adaptively retrained for each individual datum. Another limitation is that the computational load is large compared with the conventional approaches to estimate the HR and EE using a wrist band-typed photoplethysmogram (PPG) sensor (deep learning model size: approximately 70 mb, testing time: a few seconds). However, the existing HR and EE measurement devices have disadvantages when worn on a wrist, as some users feel uncomfortable to wear. In addition, they are too sensitive to noise, resulting in poor SNR. On the other hand, the proposed shoe sensor could be more natural for use to wear compared to the wrist-typed sensor.

For the future research, it would be possible to improve the generalization performance using more diverse datasets and adding personal information (gender, BMI, foot size, etc.) to the model input. It will also include the investigation of the sensor-specific functions corresponding to the variations in HR and EE values.

Author Contributions: Conceptualization and methodology, J.R. and H.E.; validation and software, H.E.; formal analysis, J.R., H.E. and S.B.; investigation, J.R. and S.L.; data curation, J.R., H.E. and Y.S.H.; writing of the original draft preparation, H.E.; writing—review and editing, S.K. and C.P.; visualization, H.E.; supervision, C.P.; project administration, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) by the South Korean government (NRF-2017R1A5A1015596), the Research Grant of Kwangwoon University in 2021, and the Ministry of Trade, Industry and Energy (MOTIE), Korea as “Development of shoes and contents solution for walk and running tracking analysis (20002862)”.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Institutional Review Board (IRB No. P01-201908-11-002).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical reasons.

Conflicts of Interest: The authors declare no conflict of interest. The funding agencies had no role in the design of the study; in the collection, analyses, and interpretation of data; in the writing of the manuscript; and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
EE	Energy expenditure
PA	Physical activity
CVD	Cardiovascular disease
HRV	Heart rate variability
HR	Heart rate
ECG	Electrocardiogram
PPG	Photoplethysmogram
ANN	Artificial neural network
CNN	Convolutional neural network
GRU	Gated recurrent unit
GAP	Global average pooling
Bi-GRU	Bidirectional gated recurrent unit
FC	Fully connected
MSE	Mean squared error
LOSO	Leave-one-subject-out
RMSE	Root-mean-square error
MAE	Mean absolute error
R^2	Coefficient of determination
LOA	Limit of agreement
MAPE	Mean absolute percent error
ANOVA	Analysis of variance

References

- Park, S.; Jayaraman, S. Enhancing the quality of life through wearable technology. *IEEE Eng. Med. Biol. Mag.* **2003**, *22*, 41–48. [[CrossRef](#)] [[PubMed](#)]
- Seneviratne, S.; Hu, Y.; Nguyen, T.; Lan, G.; Khalifa, S.; Thilakarathna, K.; Hassan, M.; Seneviratne, A. A survey of wearable devices and challenges. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 2573–2620. [[CrossRef](#)]
- Grand View Research, Inc., Wearable Technology Market Size, Share & Trends Analysis Report By Product (Wrist-wear, Eye-wear & Head-wear, Foot-wear, Neck-wear, Body-wear), By Application, By Region, Furthermore, Segment Forecasts, 2020–2027y. Available online: <https://www.grandviewresearch.com/industry-analysis/wearable-technology-market> (accessed on 27 May 2021).
- Ion, M.; Dinulescu, S.; Firtat, B.; Savin, M.; Ionescu, O.N.; Moldovan, C. Design and Fabrication of a New Wearable Pressure Sensor for Blood Pressure Monitoring. *Sensors* **2021**, *21*, 2075. [[CrossRef](#)]
- Bonato, P. Advances in Wearable Technology and Applications in Physical Medicine and Rehabilitation; *J. Neuroeng. Rehabil.* **2005**, *2*, 2. [[CrossRef](#)]
- Lin, F.; Wang, A.; Song, C.; Xu, W.; Li, Z.; Li, Q. A comparative study of smart insole on real-world step count. In Proceedings of the 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 12 December 2015; pp. 1–6.
- Fulk, G.D.; Sazonov, E. Using sensors to measure activity in people with stroke. *Top. Stroke Rehabil.* **2011**, *18*, 746–757. [[CrossRef](#)] [[PubMed](#)]
- Fulk, G.D.; Edgar, S.R.; Bierwirth, R.; Hart, P.; Lopez-Meyer, P.; Sazonov, E. Identifying activity levels and steps in people with stroke using a novel shoe-based sensor. *J. Neurol. Phys. Ther.* **2012**, *36*, 100. [[CrossRef](#)]
- Cho, H. Design and implementation of a lightweight smart insole for gait analysis. In Proceedings of the 2017 IEEE Trust-com/BigDataSE/ICISS, Sydney, NSW, Australia, 1–4 August 2017; pp. 792–797.
- Benocci, M.; Rocchi, L.; Farella, E.; Chiari, L.; Benini, L. A wireless system for gait and posture analysis based on pressure insoles and Inertial Measurement Units. In Proceedings of the 2009 3rd International Conference on Pervasive Computing Technologies for Healthcare, London, UK, 1–3 April 2009; pp. 1–6.
- Lin, F.; Wang, A.; Zhuang, Y.; Tomita, M.R.; Xu, W. Smart insole: A wearable sensor device for unobtrusive gait monitoring in daily life. *IEEE Trans. Ind. Inform.* **2016**, *12*, 2281–2291. [[CrossRef](#)]
- Truong, P.H.; Lee, J.; Kwon, A.R.; Jeong, G.M. Stride counting in human walking and walking distance estimation using insole sensors. *Sensors* **2016**, *16*, 823. [[CrossRef](#)]
- Ngueleu, A.M.; Blanchette, A.K.; Bouyer, L.; Maltais, D.; McFadyen, B.J.; Moffet, H.; Batcho, C.S. Design and accuracy of an instrumented insole using pressure sensors for step count. *Sensors* **2019**, *19*, 984. [[CrossRef](#)]
- Sazonova, N.; Browning, R.C.; Sazonov, E. Accurate prediction of energy expenditure using a shoe-based activity monitor. *Med. Sci. Sports Exerc.* **2011**, *43*, 1312–1321. [[CrossRef](#)]

15. U.S. Department of Health and Human Services. *The Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity*; US Government Printing Office: Rockville, MD, USA, 2001
16. Blair, S.N.; Kohl, H.W.; Gordon, N.F.; Paffenbarger, R.S., Jr. How much physical activity is good for health? *Annu. Rev. Public Health* **1992**, *13*, 99–126. [[CrossRef](#)]
17. Manley, A.F. *Physical Activity and Health: A Report of the Surgeon General*; U.S. Department of Health and Human Services: Atlanta, GA, USA, 1996.
18. Crouter, S.E.; Clowers, K.G.; Bassett, D.R., Jr. A novel method for using accelerometer data to predict energy expenditure. *J. Appl. Physiol.* **2006**, *100*, 1324–1331. [[CrossRef](#)]
19. Nielson, R.; Vehrs, P.R.; Fellingham, G.W.; Hager, R.; Prusak, K.A. Step counts and energy expenditure as estimated by pedometry during treadmill walking at different stride frequencies. *J. Phys. Act. Health* **2011**, *8*, 1004–1013. [[CrossRef](#)]
20. Chen, K.Y.; Sun, M. Improving energy expenditure estimation by using a triaxial accelerometer. *J. Appl. Physiol.* **1997**, *83*, 2112–2122. [[CrossRef](#)]
21. Staudenmayer, J.; Pober, D.; Crouter, S.; Bassett, D.; Freedson, P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J. Appl. Physiol.* **2009**, *107*, 1300–1307. [[CrossRef](#)]
22. Voleno, M.; Redmond, S.J.; Cerutti, S.; Lovell, N.H. Energy expenditure estimation using triaxial accelerometry and barometric pressure measurement. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 5185–5188.
23. Vathsangam, H.; Emken, A.; Schroeder, E.T.; Spruijt-Metz, D.; Sukhatme, G.S. Determining energy expenditure from treadmill walking using hip-worn inertial sensors: An experimental study. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2804–2815. [[CrossRef](#)] [[PubMed](#)]
24. World Health Organization. *Prevention of Cardiovascular Disease*; World Health Organization: Geneva, Switzerland, 2007.
25. Thayer, J.F.; Yamamoto, S.S.; Brosschot, J.F. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *Int. J. Cardiol.* **2010**, *141*, 122–131. [[CrossRef](#)]
26. Lee, E.S.; Lee, J.S.; Joo, M.C.; Kim, J.H.; Noh, S.E. Accuracy of heart rate measurement using smartphones during treadmill exercise in male patients with ischemic heart disease. *Ann. Rehabil. Med.* **2017**, *41*, 129. [[CrossRef](#)]
27. Lee, H.; Lee, H.; Whang, M. An enhanced method to estimate heart rate from seismocardiography via ensemble averaging of body movements at six degrees of freedom. *Sensors* **2018**, *18*, 238. [[CrossRef](#)] [[PubMed](#)]
28. Shashikumar, S.P.; Shah, A.J.; Li, Q.; Clifford, G.D.; Nemati, S. A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. In Proceedings of the 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Orlando, FL, USA, 16–19 February 2017; pp. 141–144.
29. Ravi, D.; Wong, C.; Lo, B.; Yang, G.Z. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 56–64. [[CrossRef](#)] [[PubMed](#)]
30. Gotlibovych, I.; Crawford, S.; Goyal, D.; Liu, J.; Kerem, Y.; Benaron, D.; Yilmaz, D.; Marcus, G.; Li, Y. End-to-end deep learning from raw sensor data: Atrial fibrillation detection using wearables. *arXiv* **2018**, arXiv:1807.10707.
31. Faust, O.; Hagiwara, Y.; Hong, T.J.; Lih, O.S.; Acharya, U.R. Deep learning for healthcare applications based on physiological signals: A review. *Comput. Methods Progr. Biomed.* **2018**, *161*, 1–13. [[CrossRef](#)]
32. Yin, Z.; Zhao, M.; Wang, Y.; Yang, J.; Zhang, J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Comput. Methods Progr. Biomed.* **2017**, *140*, 93–110. [[CrossRef](#)] [[PubMed](#)]
33. Zhu, J.; Pande, A.; Mohapatra, P.; Han, J.J. Using deep learning for energy expenditure estimation with wearable sensors. In Proceedings of the 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), Boston, MA, USA, 14–17 October 2015; pp. 501–506.
34. Yao, X.; Li, X.; Ye, Q.; Huang, Y.; Cheng, Q.; Zhang, G.Q. A robust deep learning approach for automatic classification of seizures against non-seizures. *arXiv* **2018**, arXiv:1812.06562.
35. Fu, L.; Lu, B.; Nie, B.; Peng, Z.; Liu, H.; Pi, X. Hybrid Network with Attention Mechanism for Detection and Location of Myocardial Infarction Based on 12-Lead Electrocardiogram Signals. *Sensors* **2020**, *20*, 1020. [[CrossRef](#)] [[PubMed](#)]
36. Yao, Q.; Wang, R.; Fan, X.; Liu, J.; Li, Y. Multi-class Arrhythmia detection from 12 to lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network. *Inf. Fusion* **2020**, *53*, 174–182. [[CrossRef](#)]
37. Gunning, D. *Explainable Artificial Intelligence (xai)*; Defense Advanced Research Projects Agency (DARPA): Arlington, VA, USA, 2017; Volume 2.
38. Busseti, E.; Osband, I.; Wong, S. *Deep Learning for Time Series Modeling*; Technical Report; Stanford University: Stanford, CA, USA, 2012; pp. 1–5.
39. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
40. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
41. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
42. Han, J.; Moraga, C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In Proceedings of the International Workshop on Artificial Neural Networks, Torremolinos, Spain, 7–9 June 1995; pp. 195–201.

43. Zhang, J.; Lu, C.; Li, X.; Kim, H.J.; Wang, J. A full convolutional network based on DenseNet for remote sensing scene classification. *Math. Biosci. Eng.* **2019**, *16*, 3345–3367. [[CrossRef](#)]
44. Zhang, K.; Guo, Y.; Wang, X.; Yuan, J.; Ding, Q. Multiple feature reweight densenet for image classification. *IEEE Access* **2019**, *7*, 9872–9880. [[CrossRef](#)]
45. Tao, Y.; Xu, M.; Lu, Z.; Zhong, Y. DenseNet-based depth-width double reinforced deep learning neural network for high-resolution remote sensing image per-pixel classification. *Remote Sens.* **2018**, *10*, 779. [[CrossRef](#)]
46. Tong, W.; Chen, W.; Han, W.; Li, X.; Wang, L. Channel-attention-based DenseNet network for remote sensing image scene classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4121–4132. [[CrossRef](#)]
47. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
48. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
49. Altman, D.G.; Bland, J.M. Measurement in medicine: the analysis of method comparison studies. *J. R. Stat. Soc. Ser. D Stat.* **1983**, *32*, 307–317. [[CrossRef](#)]
50. Nelson, B.W.; Allen, N.B. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. *JMIR mHealth uHealth* **2019**, *7*, e10828. [[CrossRef](#)]
51. Javed, A.R.; Sarwar, M.U.; Khan, S.; Iwendi, C.; Mittal, M.; Kumar, N. Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition. *Sensors* **2020**, *20*, 2216. [[CrossRef](#)] [[PubMed](#)]
52. Smith, M.P.; Horsch, A.; Standl, M.; Heinrich, J.; Schulz, H. Uni-and triaxial accelerometric signals agree during daily routine, but show differences between sports. *Sci. Rep.* **2018**, *8*, 1–8. [[CrossRef](#)] [[PubMed](#)]
53. Ceesay, S.M.; Prentice, A.M.; Day, K.C.; Murgatroyd, P.R.; Goldberg, G.R.; Scott, W.; Spurr, G. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *Br. J. Nutr.* **1989**, *61*, 175–186. [[CrossRef](#)]