DeepHAR: A Deep Feed-Forward Neural Network Algorithm for Smart Insole-based Human Activity Recognition

Document Version
Publisher's PDF, also known as Version of record

ORIGINAL ARTICLE

DeepHAR: a deep feed-forward neural network algorithm for smart insole-based human activity recognition

Luigi D'Arco[1] · Haiying Wang[1] · Huiru Zheng[1]

Abstract

Health monitoring, rehabilitation, and fitness are just a few domains where human activity recognition can be applied. In this study, a deep learning approach has been proposed to recognise ambulation and fitness activities from data collected by five participants using smart insoles. Smart insoles, consisting of pressure and inertial sensors, allowed for seamless data collection while minimising user discomfort, laying the baseline for the development of a monitoring and/or rehabilitation system for everyday life. The key objective has been to enhance the deep learning model performance through several techniques, including data segmentation with overlapping technique (2 s with 50% overlap), signal down-sampling by averaging contiguous samples, and a cost-sensitive re-weighting strategy for the loss function for handling the imbalanced dataset. The proposed solution achieved an Accuracy and F1-Score of 98.56% and 98.57%, respectively. The Sitting activities obtained the highest degree of recognition, closely followed by the Spinning Bike class, but fitness activities were recognised at a higher rate than ambulation activities. A comparative analysis was carried out both to determine the impact that pre-processing had on the proposed core architecture and to compare the proposed solution with existing state-of-the-art solutions. The results, in addition to demonstrating how deep learning solutions outperformed those of shallow machine learning, showed that in our solution the use of data pre-processing increased performance by about 2%, optimising the handling of the imbalanced dataset and allowing a relatively simple network to outperform more complex networks, reducing the computational impact required for such applications.

Keywords Activity recognition · Deep learning · Smart insole · Imbalance classification

1 Introduction

Human activity recognition (HAR) can be used to monitor user's behaviours, analyse them, and consequently assist the user in his/her daily life or provide histories on the activities to specialists for evaluation. The applications of HAR include health monitoring [1, 2], rehabilitation [3], fitness [4], home automation [5], and safety [6].

The pioneering activity recognition approach has been based on the analysis of visual data, including both images and videos [7]. Considering the dynamism with which a person performs an activity during his/her daily living, multiple challenges can be found in the use of vision-based solutions, including viewpoint variations, occlusions, cluttered backgrounds, different illumination conditions, and privacy concerns [8, 9]. As a result, alternative solutions have been studied in recent years, such as those based on the use of sensors, which can be positioned in the environment surrounding the user or directly worn by the same [10]. Wearable-based solutions have steadily become the centre of research due to their extensive computational power, minimum encumbrance for the user, and low costs. Wearable technologies include smartphones, smart watches, smart clothes, and other specifically designed devices. Generally, fusing multiple heterogeneous sensors, which measure the same physical phenomenon, can

✉ Huiru Zheng
  h.zheng@ulster.ac.uk

  Luigi D'Arco
  darco-l@ulster.ac.uk

  Haiying Wang
  hy.wang@ulster.ac.uk

1 School of Computing, Ulster University, York st, Belfast BT15 1ED, Northern Ireland, UK

increase the variability and the insight of the information that can be better exploited for classification purposes [11]; nevertheless, this can cause discomfort for the user and can increase exponentially the cost of the solutions. In this regard, smart insoles have attracted significant attention recently, since they can embed multiple sensors and seamlessly integrate into users' daily lives.

Smart insoles are specialised inserts that can be placed inside a pair of shoes to collect and monitor various forms of data about the user's foot activity and movements. They are equipped with sensors and other electronic components that allow them to gather and transmit data wirelessly to a smartphone or other device. Generally, pressure sensors and inertial sensors are the smart insoles embedded sensors preferred in literature, but not limited to. Pressure sensors allow measuring the force exerted by the foot while carrying activities and can be classified into piezoresistive sensors, capacitive sensors, and optical sensors. Inertial sensors, also known as inertial measurement units (IMUs), are devices that are used to measure and track the acceleration, orientation, and angular velocity of an object in three-dimensional space. They typically consist of a combination of accelerometers, gyroscopes, and magnetometers. The sensors used in the state-of-the-art solutions vary according to the needs of the study, e.g. Chen et al. [12] proposed a smart insole composed of a pressure array (up to 96 pressure sensors distributed over the insole), a triaxial accelerometer, and a triaxial gyroscope, while Aznar-Gimeno et al. [13] designed a smart insole composed of 16 piezoelectric sensors, a triaxial accelerometer, and a temperature sensor.

Although the capabilities of such devices are improving and their performance is rising, HAR is still a complicated task. Each activity, by its nature, is difficult to recognise because it can be influenced by a variety of situations, and even the same person can execute the same activity in two different ways depending on the circumstances [14].

Several approaches can be identified in the literature to process and analyse the amount of data produced by the various sensors embedded inside the smart insoles, including threshold algorithms, and machine learning solutions. Threshold-based algorithms allow the classification of activities based on predefined rules determined by experts, identifying ranges in which each activity falls. Machine learning-based algorithms, instead, analyse a set of data collected a priori from volunteers and attempt to identify patterns in those data that can be used to generalise the problem and classify the activities without human intervention. Since threshold-based algorithms require manual adjustments to their parameters or rules when the data or circumstances change, which can be time-consuming and may require expert knowledge, machine learning solutions are the most preferable. Independently

from the machine learning algorithm chosen, the goal of enhancing the recognition accuracy is usually allocated to the extraction of features from raw data, which can be broadly classified into feature-based and feature-learning approaches [15]. In feature-based approaches, the features are extracted by experts using heuristic-based methods [16], whereas, in feature-learning approaches, the salience information is extracted automatically by the algorithm chosen [17], which is commonly a deep learning algorithm. Deep learning algorithms are inspired by the structure and function of the human brain. They are composed of multiple layers of interconnected nodes, in which each layer extracts increasingly complex features from the raw data. These features are then used to make predictions or decisions based on the data inputs. Overall, deep learning is a powerful tool that has been used effectively in the field of human activity recognition using wearable sensors, due to its ability to learn and adapt to new data and its versatility in handling a wide range of data types and structures. Although this type of solution is very effective, it requires a large amount of data samples for training and evaluation, which when coupled with the challenges of obtaining activity data results in issues like class imbalance. Due to highly demanding training and evaluation and large memory requirements, they are computationally intensive, making them difficult to integrate into portable devices and provide real-time responses. Furthermore, when using small datasets and/or complex architecture, these systems are vulnerable to overfitting [18].

The aim of this paper was to propose a deep learning approach for the recognition of ambulation and fitness activities using smart insoles, which can potentially be integrated into daily life scenarios for physical activity monitoring and/or rehabilitation. The smart insoles consist of eight pressure sensors and a nine-degree inertial measurement unit (IMU), consisting of an accelerometer, gyroscope and magnetometer. To facilitate and simplify the data acquisition process, a mobile application has been developed, which provides data collection, visualisation and archiving functions which, combined with a cloud server, allow recognition of the activities. A deep feed-forward neural network (henceforth referred to as DeepHAR) has been implemented for the recognition of activities. The key objective of this work was to prove that the performance of such an architecture, despite it being a relatively simpler architecture, with adequate pre-processing can exceed more complex solutions such as convolutional neural networks (CNN), preventing the overfitting and reducing the computation costs of the solution limiting the number of hyperparameters and layers in the architecture. To enhance the solution performance, a time-windowing technique with overlap between contiguous segments and a down-sampling technique for denoising

raw sensor signals have been involved. Furthermore, to solve the problem of imbalanced classes, the architecture has been equipped with a loss function that considers the weights calculated for each class.

The rest of the paper is organised as follows: state-of-the-art solutions are introduced in Sect. 2, and hardware details and the methodology are presented in Sect. 3, followed by the findings discussed in Sect. 4. Finally, the paper is concluded by a summary in Sect. 5 and future work in Sect. 6.

2 Related work

Human activity recognition is one of the most important tasks in pervasive computing. Over the years, efforts have been made to enhance and optimise the proposed solutions, using different technologies and devices. Adopting wearable devices-based solutions has made it possible to develop seamless solutions and to reduce encumbrance for the user.

Over the past few years, smart insoles have become increasingly popular due to their noticeable benefits and minimal user inconvenience. They have been acknowledged as healthcare devices, and there are numerous commercially available solutions, including the OpenGo system from Moticon ReGo AG [19], the Smart Footwear from IEE Luxembourg S.A. [20], and the Neurogait insoles from Salted Ltd. [21]. The major differences between them are the types and number of sensors used.

In terms of the HAR algorithms used, the desire for maximum optimisation has resulted in a heterogeneous set of algorithms. The most popular solutions are those that require expert supervision, such as customised threshold algorithms and machine learning algorithms, in which feature extraction techniques are critical for performance optimisation [22].

Moufawad el Achkar et al. [23] proposed a solution for monitoring the risk of falls and frailty in the elderly, using instrumented shoes. A triaxial accelerometer, triaxial gyroscope, triaxial magnetometer, eight pressure sensors, and a barometer sensor were included in the instrumented shoe. The data were obtained with a sampling frequency of 200 Hz from 10 elderly people and then segmented using a window size of 5 s with 50% overlap. The HAR algorithm was a biomechanics-inspired expert-based decision tree, which analysed the locomotion or not and used the values of the sensors above thresholds to recognise the activities carried out by the person. Nine activities were included: level walking, downhill, downstairs, uphill, upstairs, sitting, standing, elevator down, and elevator up. The overall *Accuracy* of the system was 97.41%, with low sensitivity (79%) for the elevator up and down.

De Pinho et al. [24] exhibited a six-activity classes machine learning HAR classifier using a foot-based wearable device. The wearable devices consisted of two components: a smart insole with six pressure sensors and a microcontroller that managed an inertial measurement system comprised of an accelerometer, gyroscope, magnetometer and barometer. Eleven participants were included in the study, which performed different activities in a controlled environment, including walking (straight, slope up, slope down), and ascending and descending stairs. The sampling frequency for the data collection was set to 10 Hz, and the data were segmented using a time windowing of 0.3 s. Initially, a set of 100 features were selected, comprising mean, standard deviation, variance, minimum, maximum, and average value; however, after feature selection using Hall's algorithm the features were reduced to 12. The random forest was used as classified and the training and testing phases were carried out involving a leave-one-out cross-validation strategy. The RF reached an overall *Accuracy* of 93.34%.

Sazonov et al. [25] described a shoe-based wearable sensor solution that operates with a smartphone to recognise various physical activities in real time and estimate energy expenditure. The smart shoes presented embedded five pressure sensors and an accelerometer. Four activities were included in the study: sitting, standing, walking/logging, and cycling, collected from 19 participants which wore the smart shoes for almost four hours. The data were collected using a sampling frequency of 400 Hz, but then to remove the possible noise they opted to average the 16 consecutive samples, reducing the actual sample frequency to 25 Hz. The data were segmented using a time windowing of 2 s, and the following features were extracted for each sensor: mean, entropy, and standard deviation. Three algorithms were used for activity classification, the support vector machine (SVM), the multi-layer perceptron (MLP), and the multinomial logistic discrimination (MLD). The SVM reached the highest performance with an *Accuracy* and *F1-Score* of 97.9% and 98.4%; nevertheless, the MLP and the MLD reached almost comparable results, but reducing the running time and the memory requirements by a factor of 10^3.

These solutions were based on data processing and in particular on the extraction of features. However, these characteristics were heuristically chosen, which could lead to poor results when analysing new data. Feature selection techniques can be used to reduce irrelevant features [26], but they still use the initially determined set of features. As a result, algorithms that allow the processing of raw data, such as deep learning models, have become increasingly popular in recent years.

Pham et al. [27] presented a convolutional neural network (CNN) for identifying physical activities such as running, walking, standing, jumping, kicking, and cycling. A 3D accelerometer sensor built into a pair of shoes was employed. The data were captured at a sampling rate of 50 Hz and segmented using a 2-s sliding window technique with a 50% overlap between two consecutive windows. The study involved ten participants who were given 10 to 30 min to complete each exercise. The CNN was built by reserving a CNN for each sensor signal in input, and then, the CNNs results are concatenated in a fully connected network for the activity prediction. The CNN was tested using a tenfold cross-validation method, which yielded an average *Precision* and *Recall* of 93.41% and 93.16%, respectively.

Wang et al. [28] proposed a one-dimensional convolutional neural network (CNN) for the recognition of activities of daily living (ADLs) against falls. The ADLs used were: laying on the bed, bowing, walking, jogging, and laying down. Two sensors were embedded into smart insoles, a triaxial accelerometer and a triaxial gyroscope. To train the model, the data from 10 healthy volunteers were collected, and to isolate each activity, the data were segmented into six-second time windows. Falls were recognised with an overall *Accuracy* of 98.61% and exhibited high sensitivity and specificity, 97.92% and 99.58%, respectively. In addition, the results showed that the walking and jogging activities were detected with an *Accuracy* of 100%.

Paydarfar et al. [29] developed a HAR system using piezoresistor-based instrumented shoes and a recurrent neural network (RNN). A pair of sneakers with an integrated microcontroller and three piezoresistor sensors at the calcaneus, metatarsals, and phalanges made up the hardware. The experiment involved 20 healthy people. Each participant performed different activities, including walking, standing, balancing on the left foot, balancing on the right foot, toe-up, and ascending stairs. Each task was performed for 45 to 120 s. The data were sampled at a frequency of 50 Hz and successively segmented into one-second slices, but each slice differs from the preceding by only one time-step. The system obtained an overall *Accuracy* of 87%.

In our previous study [30], an artificial neural network (ANN) was implemented for the recognition of ambulation activities. Three volunteers were involved in the study and were asked to wear a pair of smart insoles and complete a series of activities from a predefined set, including downstairs, sit to stand, sitting, standing, upstairs, and walking (slow, normal, and fast). Given the unbalanced nature of the dataset used, a data over-sampling technique was used, the SMOTE, which created synthetic data to level the number of samples for each class. The ANN developed consisted of two fully connected layers preceded by a flattened layer to squeeze the input data. The results of this preliminary study showed that the performance of the classifier was mainly influenced by the over-sampling technique, which in order to balance the number of samples for each class created several synthetic data, with consequent reduction of the variance and entropy in the data.

Considering the advances of soft-computing solutions [31, 32] in real-life applications and the promising results achieved by deep learning algorithms in the processing of sensor data, this study has been established to overcome the limitations encountered in our previous study as well as the limitations that arose during the literature review. The main challenge identified has been the treatment of the imbalanced dataset for the training of deep learning models. Furthermore, the solution involving deep learning has shown complex architectures and extensive hyperparameters, which have significant computational time and expensive costs when considering a real-life application. For this reason, it has been investigated in the study how simpler neural networks, such as a feed-forward neural network, can achieve results comparable if not superior to more complex networks when data from the smart insole are used and data pre-processing techniques are applied. The choice of architecture, associated with a search for the minimum number of layers that can optimise the classification, provides a reduction in computational costs, which associated with the extension of the activities involved to both ambulation and fitness, lays the foundations for the use of the solution in real-time scenarios, such as monitoring or rehabilitation of an individual.

3 Materials and methods

In this study, a smart insole-based human activity recognition (HAR) system is proposed. Figure 1 shows the overall architecture, which consists of a pair of smart insoles, a mobile application, called eZiGait, and a could server.

3.1 Measurement set-up/sensing elements

In this study, the ActiSense Kit (IEE Luxembourg S.A.) was used as the only device for the human activity recognition (HAR) system. The ActiSense kit includes two IEE Smart Foot Sensors and two ActiSense electronics (ActiSense ECU), shown in Fig. 2. The IEE Smart Foot Sensors are composed of eight individual high dynamic pressure cells, which are located at the point where the impact foot-to-ground is higher based on a finite element (FE) analysis and extensive testing and validation, as shown in Fig. 2a. The ActiSense ECU, as shown in Fig. 2b,
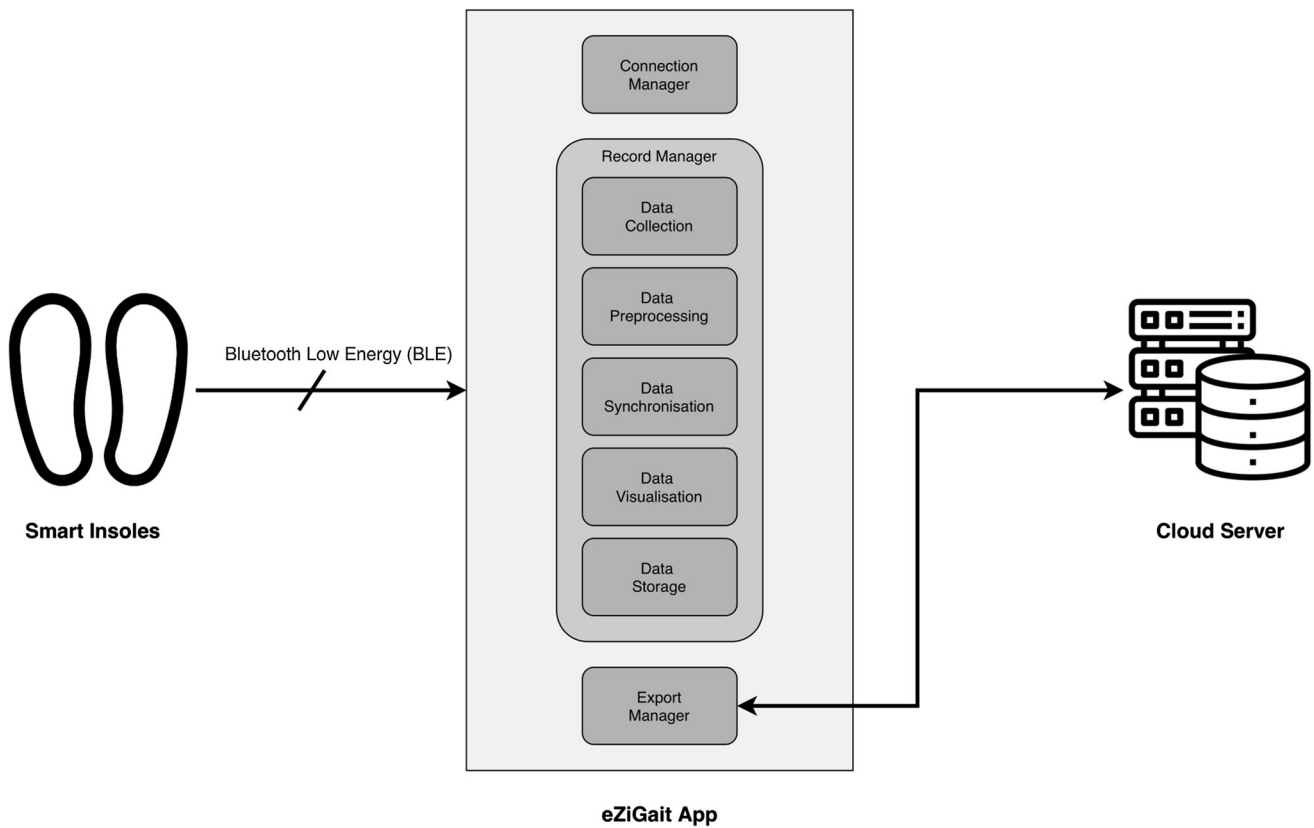
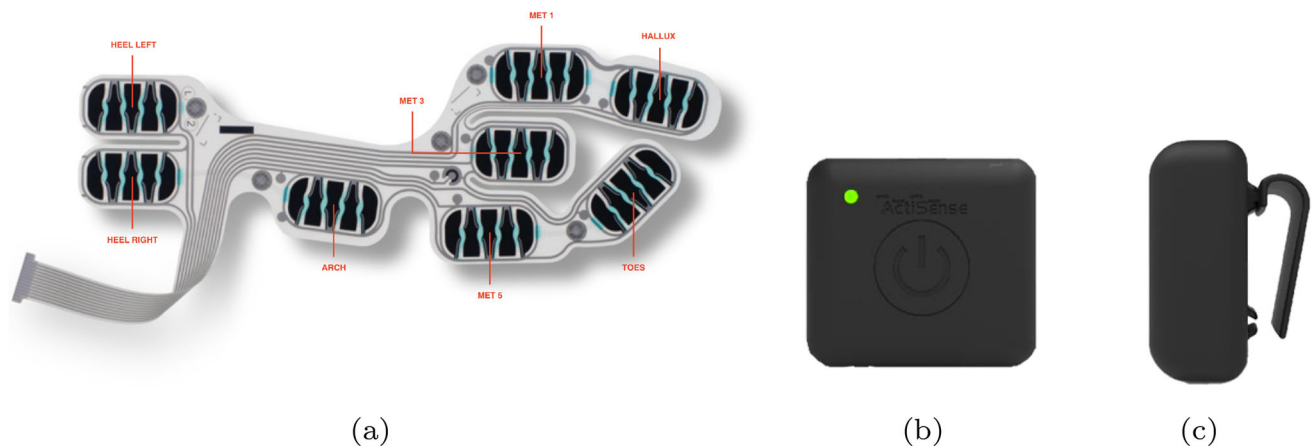**Fig. 1** Overall architecture of the proposed HAR system

(a) (b) (c)

Fig. 2 ActiSense Kit, IEE Luxembourg S.A. **a** IEE Smart Foot Sensor and **b** IEE ActiSense ECU—front **c** IEE ActiSense ECU—side

is the electronic unit of the kit. It incorporates multiple inertial measurement unit (IMU) sensors, including a triaxial accelerometer (range: ± 8 G), triaxial gyroscope (range: ± 1000 DPS), and a triaxial magnetometer (range: ± 4912 µT), providing the user with a nine-degree-of-freedom (DOF) system. Furthermore, a temperature sensor is included in the unit, but currently, it is not used in this study. The data collected can be transferred from the ActiSense ECU via the Bluetooth Low Energy (BLE) protocol to a smartphone, or stored locally on flash memory. The ActiSense ECU is attached to the side of the shoe using the hook provided, as shown in Fig. 2c.

3.2 Mobile application

The mobile application (i.e. eZiGait) has been developed from the prototype presented by McCalmont et al. [33] for data collection and visualisation. It is the central

component of the proposed system architecture, as shown in Fig. 1, as it handles the connection with the smart insoles through BLE and gathers data. Furthermore, it is connected to the cloud server for saving data and retrieving activity recognition results. The app has been developed using stackable modules, called managers, for allowing the inclusion of new modules as the requirements change. The management of data collection is delegated to the record manager module which starts a data stream from the insoles and processes it. The raw data collected are converted into data ready for pre-processing. The data from the two insoles are then synchronised with each other by coupling the samples coming from the same timestamp. Furthermore, during the data collection, data visualisation functions allow the user to view the data in real time. The last phase is to save the data, which are preserved locally, on the smartphone itself, and on cloud storage via the HAR manager, which, in turn, provides the user with a classification of the activity undertaken.

3.3 Data collection

Human activities can be grouped into seven main categories: ambulation, transportation, phone usage, daily activities, fitness, military, and upper body activities [34]. Smart insoles have been applied for the detection of ambulation activities, daily activities, and fitness exercises; however, it has been proven that they cannot be used alone to detect activities that involve only the upper body [35, 36]. For this reason, a set of activities have been defined comprising ambulation and fitness activities. The activities included are: Walking (Slow, Normal, Fast, Free), Sitting, Standing, Ascending Stairs, Descending Stairs, Cross Trainer, Sit to Stand, Spinning Bike, Standing, and Free Stretch. The descriptions of the activities and their collection modalities used in this study are summarised in Table 1.

Five participants (age: 25–55 years, weight: 48.0−75.0 kg, and height: 165.0−180.0 cm), comprising European and Asian people, with no reported lower limb injuries were recruited for this study. All the participants were provided with an ActiSense Kit according to their shoe size. All the data were collected using the eZiGait App, using a sampling frequency of 200 Hz.

Each participant had the freedom to choose which activity to perform among those designated. In total, 178 min of recordings were collected.

3.4 Data pre-processing

The raw data collected by the sensors can have imperfections which can in turn affect the performance of the solution. Hence, enhancing the representation of the input

can improve the final prediction outcome. Recently, multiple data pre-processing techniques have been adopted in the literature to enhance the accuracy outcome of solutions in several scopes [37–39]. The data collected from the smart insoles combined multi-modal data information, since they combined pressure and inertial data, which are all in the form of continuous data. In this study, in addition to the normalisation technique, which converts the input data into a range between 0 and 1, other techniques were introduced to improve data representation, including the interpolation technique for handling missing data, downsampling technique by averaging contiguous samples for noise reduction, and time-windowing data segmentation. Furthermore, the weight associated with each class according to the number of samples is calculated to avoid training bias towards the majority classes.

3.4.1 Handling missing data

Generally, statistical models are designed with the assumption that no observations are missing when processing the data. For this reason, dealing with missing data is crucial to prevent failure and unexpected model outcomes. Three basic categories of missingness may be identified: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MCAR occurs when missing observations are dependent on both observed and unobserved measurements. MAR occurs when the likelihood of missing observations is only connected to observable data. MNAR occurs when missing observations are not reliant on either observed values or unseen values [40]. In the literature, there are multiple approaches for solving this problem, including deleting incomplete observations and replacing the missing values with an estimate based on other available information, also known as imputation [41]. Although the deletion of missing data is the most common, it has significant drawbacks, including decreasing statistical power due to the smaller number of samples and the potential to change the representation of the population by favouring one subgroup over another. The imputation, on the other way, substitutes missing values by using statistical measurements, such as mean or median, interpolation using existing information, or using a model-based approach such as linear regression or stochastic regression. Statistical approaches, however, reduce variability and the correlation within and between variables, whereas model-based solutions can create better estimations than the true values or their performance can be poor due to the non-relationship between missing and observed values.

In this study, the type of missingness identified is of the MCAR type, as the missing observations are mainly due to random faults in the data transmission from the device. For

Table 1 Description of the set of activities and their collection modalities involved in this study

Name	Definition
Cross trainer	Activity carried out on gym equipment, on which the person can exercise his arms and legs by simulating the activity of walking, running or climbing stairs. The speed can vary from 50 to 70 rpm
Downstairs	Descend the stairs, placing one foot on the step and the other on the next
Free walking	Locomotion in a free environment. Walking can occur in all directions and along curved lines
Sit to stand	Transition activity between sitting and standing. It starts when the person's buttocks leave the seat of the chair or bed, and it end's when the person's feet support the person's body weight without any other movements in the legs
Sitting	Person's buttocks are on the seat of the chair or bed. Sitting can include movements in the upper body
Spinning bike	Activity carried out on a stationary bike. The speed can vary from 50 to 100 rpm
Standing	Both feet supporting the person's body weight, with no feet movement. Standing can include movements in the upper body
Stretch	Activity in which a specific muscle or tendon is deliberately flexed or stretched in order to improve the muscle's felt elasticity and achieve comfortable muscle tone
Upstairs	Ascending the stairs, placing one foot on the step and the other on the next
Walking (Slow, Normal, Fast)	Locomotion on a treadmill with different speeds. Slow, Normal and Fast have a speed of 3.6, 4.7, and 5.8 km/h, respectively

this reason, the approach chosen to deal with this problem is data interpolation using the polynomial function which lends itself particularly well to use with time series [42]. This method takes into consideration adjacent data belonging to a single time series and creates a polynomial function, which passes through the existing points and recreates the missing points within the time series.

3.4.2 Noise reduction

The sensors' error could increase exponentially over time, resulting in a signal completely buried in a respective noise signal. Reducing such noise is important to provide the algorithm with a clear signal for processing, as the signal can interfere with the accuracy and reliability of the same.

There are several techniques that can be used to denoise sensor signals, including low-pass filters, median filters, Kalman filters, and wavelets denoising. By attenuating components above a specific cutoff frequency, low-pass filters can be used to remove high-frequency from a signal. If the cutoff frequency is not selected carefully, however, they can also remove important information from the signal and introduce delay, which can be problematic in real-time systems. By substituting each sample in the signal with the median value of a group of nearby samples, the median filter is used to eliminate outliers from the signals; however, it can introduce delay, as it has to compute the mean values for all the samples. The Kalman filter is a type of recursive filter that can be used to estimate the state of a system in the presence of noise; however, it requires a good estimate of the system's initial state and it can be computationally expensive when large or complex systems are involved. Wavelet denoising is a technique that uses wavelets to decompose the signal into different frequency

components and removes noise from the low-frequency components while leaving the high-frequency components intact. However, the wavelet denoising is sensitive to the choice of wavelet basis and the level of decomposition used, and it can be computationally expensive, particularly for signals with a high sampling rate.

In this study, since the sampling frequency is high (200 Hz) during the data collection, the median filter has been involved to reduce noise and remove outliers from the signal. However, instead of processing all samples, the averages of 10 contiguous samples were calculated, effectively applying a down-sampling method that reduces the number of samples per activity. Down-sampling techniques through the averaging of contiguous samples have been applied widely in the literature for activity recognition. Sazonov et al. [25] and Hedge et al. [43] applied a down-sampling method that reduced the sampling frequency from 400 to 25 Hz by averaging 16 contiguous frames, as well as Merry et al. [44], which averaged five contiguous samples, hence reducing the sampling frequency from 75 to 15 Hz. In this study, 10 contiguous samples have been averaged, reducing the sampling frequency from 200 to 20 Hz. Figure 3 illustrates an example of applying the down-sampling technique on the sensor signals for noise reduction.

3.4.3 Data segmentation

Activity data collected by participants involved in this study presented different lengths, which make it difficult to analyse and classify. Therefore, determining homogeneous segments among those data is crucial, as the classification task becomes easier and more accurate, as the model can focus on a reduced amount of data and on specific aspects.

Fig. 3 Example of the down-sampling technique for denoising sensor signals, averaging ten contiguous samples in the walking activity: **a** accelerometer axis-x and **b** gyroscope axis-x
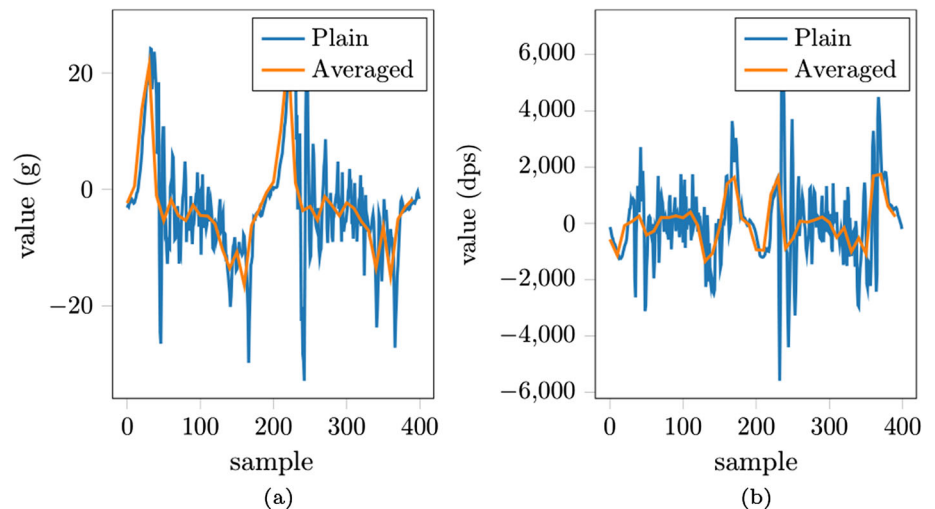
Multiple techniques can be used for the definition of the sizes of the segments and can be classified into time-based windowing, event-based windowing and dynamic windowing [45]. Time-based windowing allows data collected to be divided into fixed segments of equal size. Event-based windowing allows dividing the data according to a specific sensor or user events. Dynamic windowing is used when the data do not have a fixed structure and determines the segments using thresholds and rules. In both event-based and dynamic windowing, the segments can result in different sizes. Moreover, it is worth mentioning that data segmentation can be applied multiple times to create finer granularity in the data segments.

In this study, time-based windowing has been applied to segment the data collected. In the literature, multiple studies can be found on the definition of the optimal window size in time-based windowing, also known as the sliding window. Banos et al. [46] after analysing multiple window sizes determined that the window size between 1 and 2 s are those that better manage the trade-off between recognition speed and accuracy. Putra et al. [47] analysed multiple types of sliding windows with multiple datasets, recommending the 2 s window as optimal. Lee et al. [48] analysed the impact that multiple window sizes had on a CNN-based human activity recognition algorithm, determining the 2 s the size that allowed it to achieve the highest F1-Score.

Aware of the advances in the literature, in this study, the sliding window has been developed to segment the data into a 2 s window. However, considering that in time-based segmentation a major drawback can be the possibility of leaving important events outside the window, or on the border of the window, to enhance the capability and the performance of the solution and to account for activities that may occur between two segments, an overlapping of 50% between contiguous windows has been introduced.

3.5 Deep neural network HAR algorithm

In this study, a deep feed-forward neural network-based HAR algorithm (DeepHAR), was proposed and implemented. A feed-forward neural network maps an input x to a target category y by finding a mapping $f(x\|\Theta)$ such that it can approximate the classifier $f^*(x)$. It is composed of three or more layers that are interconnected, and the information x flows from the input through these layers and finally to the output y. Each layer has its own function and they are connected to each other in a chain (e.g. $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$ when it is constructed of three layers). Each layer is composed of nodes that try to mimic the human brain neurons' behaviour, by learning information from data. The nodes in consecutive layers form a bipartite graph. A node combines the elements of the input linearly with various weights (w_i) and passes the value obtained through an activation function. Hence, an arbitrary hidden layer can be represented as:

$$h^{(k+1)} = \alpha\left(b^{(k)} + W^{(k)}h^{(k)}\right), \tag{1}$$

where $W^{(k)} \in \mathbb{R}^{(N^{(k+1)} \times N^k)}$ contains all the weights, $b^{(k)}$ consists of all bias terms, and $h^{(k)}$ is the values of the previous layer. For the input layer, $h^{(0)} = x$.

The DeepHAR architecture proposed in this study is presented in Fig. 4, highlighting the input's size, the number of layers, and the activity labels used for prediction. It is composed of eight layers including input and output layers. The number of hidden layers was determined as optimal as a result of several experiments and kept to a minimum to reduce computational costs and the risk of a vanishing gradient problem [49].

The input layer is a flatten layer, which allows converting the input matrix ($x \in \mathbb{R}^{(40 \times 34)}$) into a column-wise shape to feed into the next layers. Hidden layers consist of

three pairs of fully connected layers and dropout layers. The fully connected layers (dense layer) consist of neurons with respective weights and biases and all the inputs are connected to every activation unit of the next layer, as presented in Eq. (1). The fully connected layers have 512 neurons each and include a batch normalisation, which re-centres and re-scales the data making the neurons' output (z) follow a standard normal distribution across the batch before applying the activation function. For each fully connected layer, the rectified linear unit (ReLU) activation function has been used. It decides whether or to what extent the input signal should pass, applying the following equation:

$$\text{ReLU}(z) = \max\{0, z\}. \tag{2}$$

ReLU(z) is linear for all positive input values and 0 for all negative values. The Dropout layers were introduced to prevent overfitting by dropping out units in the DeepHAR. Dropout layers are needed because neighbouring neurons begin to rely on some specialisation overtraining and if carried too far, it can result in a weak model that is overly specialised to the training data [50]. The probability of a neuron being dropped was set to 0.5.

During data pre-processing, the ground-truth labels were denoted as integers between 0 and $C - 1$, where C is the number of activities in the dataset. To allow the DeepHAR to predict a categorical output, the labels were converted to a one-hot vector $y \in \{0, 1\}^C$ to indicate the label where $y_i = 1$. Hence, the output layer consisted of two functions: a linear function and a softmax function. The linear function transformed the input x into a n-dimensional vector $z \in \mathbb{R}^n$ as:

$$z = Wx + b \tag{3}$$

where $W \in \mathbb{R}^{n \times d_{\text{in}}}$ and $b \in \mathbb{R}^n$. The softmax function, instead, normalised z into a discrete probability distribution over the classes as:

$$\hat{y}_i = \text{softmax}(z)_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}, i = 1, \ldots, n \tag{4}$$

where z_i denotes the ith element of the vector z, while $\hat{y}_i$ is the ith element of the output of the softmax function.

Basically, $\hat{y}_i$ denotes the likelihood that the input sample will be predicted with label i. Furthermore, the cross-entropy loss function has been employed to measure the difference between the ground truth and the prediction as follows:
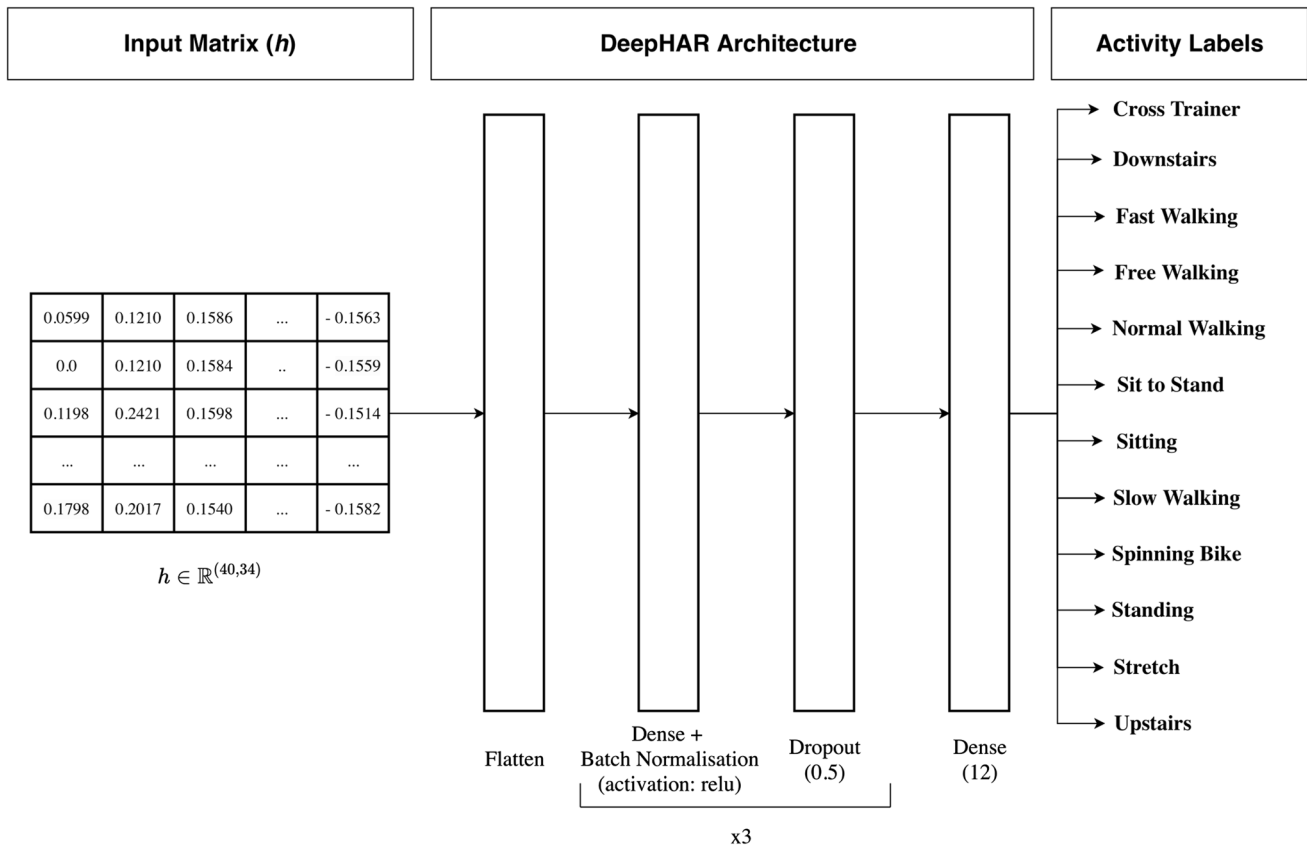

Fig. 4 Deep feed-forward neural network architecture (DeepHAR) proposed in this study for activity recognition

$$\mathcal{L}(y, \hat{y}) = -\sum_{i=0}^{n-1} y_i \log(\hat{y}_i). \tag{5}$$

3.5.1 Handling class imbalance

Machine learning algorithms assume that the data are evenly distributed across classes and no bias is present. The dataset created, however, had an uneven distribution within classes, as shown in Fig. 5. During the model training, it could occur that the predictions could have been skewed towards the majority classes. In general, two strategies could be applied [51]: re-sampling [52, 53] and cost-sensitive re-weighting [54, 55]. The re-sampling includes over-sampling (adding repetitive data) and under-sampling (removing data), and both may introduce further issues, such as the introduction of large amounts of duplicated samples making the model susceptible to overfitting in over-sampling, or the discarding of valuable samples that are important for feature learning in under-sampling. In this study, the cost-sensitive re-weighting approach was chosen, which influences the loss function by assigning higher costs to samples from the minority classes. Defined the total number of samples in the dataset as N and the number of classes in the dataset as C, the class weights must ensure that the total number of effective samples is equal to the total number of samples (N), also written as:

$$w_1 * N_1 + w_2 * N_2 + \cdots + w_C * N_C = N \tag{6}$$

where w_i is the weight for the class i, N is the total number of samples, C is the number of unique classes, and N_i is the total number of samples in the class i with $i = 1, 2, ..., C$.

Moreover, each class should have an equal number of effective samples, which can be presented as follows:

$$w_1 * N_1 = w_2 * N_2 = \cdots = w_C * N_C. \tag{7}$$

From Eqs. (6) and (7), the class weight (w_i) for the class i can be calculated as follows:

$$w_i = \frac{N}{C * N_i}. \tag{8}$$

During the training phase, the weight differences will influence the classification of the classes. The goal is to penalise the majority class by giving them a lower class weight while giving the minority class a greater weight.

The results of the computed class weight are presented in Fig. 5, with the class *Spinning Bike* having the highest number of samples (1481) and the smallest weight (0.40) assigned, and the class *Free Walking* having the lowest number of samples (90) and the highest weight (6.58).

Determined the weights that each class has on the classification, the loss function has been modified, integrating the class weights calculated using Eq. (8) into Eq. (5). Hence, the class-balanced (CB) loss function used can be written as follows:

$$CB(y, \hat{y}) = W_C \mathcal{L}(y, \hat{y}) = -\sum_{i=0}^{n-1} w_i y_i \log(\hat{y}_i) \tag{9}$$

where W_C is the vector containing all the class weights calculated and w_i is the calculated weight for the class i.

3.6 Performance assessment

To evaluate the performance of the proposed solution, the neural network was trained and tested using a cross-validation approach, which divides the dataset into equal portions and trains the model using all but one that is utilised for testing. Since the dataset used in this study is imbalanced, the validation used is a stratified cross-validation, which maintains unaltered the ratio between the

Fig. 5 Distribution of samples with relative class weights in a training set during cross-validation
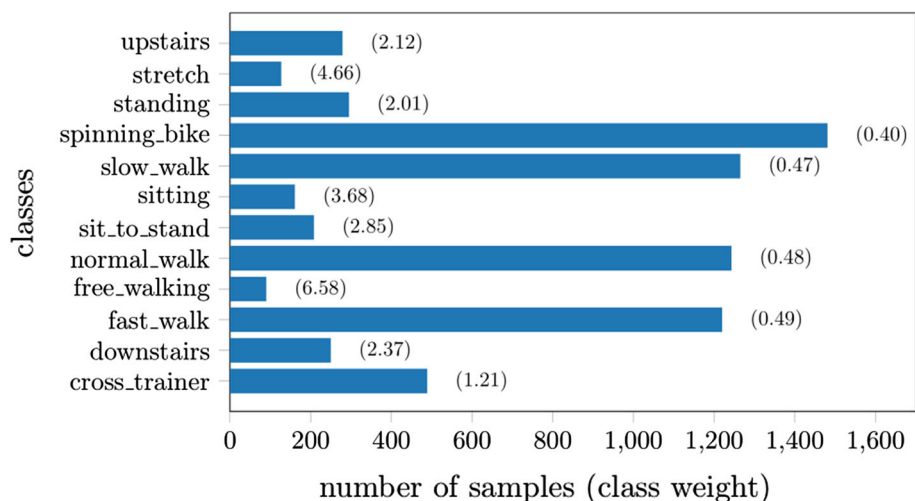
number of samples per class in the different portions. The number of portions and the number of repetitions of the evaluation were set to 10. Alternatives solutions for evaluating the model's performance are the plain k-fold cross-validation, the manual splitting in training and test sets and the leave-one-subject-out cross-validation; however, they were not considered in this study because they can be affected by the samples balance and the size of the dataset used.

The outcomes of the evaluation were utilised to create a confusion matrix, which is an error matrix that contrasts the observations estimated by the solution with the ground truth, the observations of reference. The confusion matrix was used to extract four key evaluation metrics, including Accuracy, Precision, Sensitivity, and F1-Score.

Accuracy is the measure of how often an algorithm correctly classifies data points. However, it can be affected by the balance of the dataset used and should therefore be accompanied by other metrics for a more robust evaluation. Precision is the number of true-positive predictions divided by the total number of positive predictions made by the algorithm. Sensitivity, also known as recall, is the proportion of true-positive predictions to the total number of actual positive samples. The F1-Score, which is the harmonic mean of precision and recall, provides a balance between these two metrics, giving an overall measure of the precision and robustness of the classifier. Except for accuracy, all metrics listed were calculated for each class separately and then merged using a weighted approach that took into account the number of samples for each class.

In addition, the area under the receiver operating characteristic (AUROC or AUC) has been included to evaluate the model performance, as it is more reliable in cases of an imbalanced dataset. It identifies the ability of the model to discriminate between positive and negative cases. The AUC can be calculated as the area under the ROC curve, which is, in turn, calculated as the trade-off between the true-positive rate and false-positive rate across different decision thresholds.

4 Results and discussion

The aim of this work was to develop a human activity recognition algorithm that can take advantage of the information collected by smart insoles. In this section, the results obtained will be discussed and reasonable considerations will be addressed.

The proposed algorithm, DeepHAR, was trained and tested on data collected by five participants using a stratified tenfold cross-validation, to ensure that the performance is constant across multiple experiments. An early stop technique was used for training the model, i.e. once the

model's performance was stable, the training was ended. A grid search investigation was defined for the identification of the DeepHAR's hyperparameters, which ended with the model being trained for a total of 31 epochs with a batch size of 32 samples and a learning rate of 10^{-3}.

The proposed solution has demonstrated exceptional performance, as evidenced by the outstanding results achieved. It exhibits an overall high level of Accuracy in recognising the different activities of 98.56%. The solution effectively showcases its ability to process and identify the different activities patterns in the data provided by the smart insoles and to deal effectively with the class imbalance issue as proven by the overall F1-Score and area under the curve (AUC) values, of 98.57% and 99.25%, respectively, which cannot be biased by definition by the number of samples for each class used during the testing.

The cumulative confusion matrix, given in Table 2, given by the use of stratified tenfold cross-validation allows for analysing and comprehending in detail the performance of the proposed solution in the recognition of each activity. The Sitting class achieved the highest level of performance, with 100% Precision and Sensitivity, closely followed by the Spinning Bike class, which achieved a Precision and Sensitivity of 99.82% and 100%, respectively. The worst performing classes were Downstairs (90.19% Precision and 87.92% Sensitivity) and Free Walking (92,74% Precision and 93,08% Sensitivity). The major misclassifications of the Downstairs activities are related to Upstairs activities. The two activities can result in the same pressure and acceleration patterns depending on the user who performs the activity as in both cases the foot could rest completely on the ground and the swing between one step and another is almost similar. Furthermore, the misclassification reasons can be traced back to the lack of altitude information that did not allow the algorithm to understand the direction in which the users were walking, even if a variation of that was identified. This issue could potentially be addressed in future work by incorporating a barometer, which reports altitude data. Overall, the misclassification rate between Downstairs and Upstairs is about 7% of the samples, which requires further investigation of their purity. Moreover, Downstairs activities were wrongly classified as Sit to Stand or Walking activities. The incorrect classification of the Sit to Stand in the Downstairs estimates can be associated with the change in pressure when there is a phase of oscillation between one step and another followed by a strong pressure of the foot that first touches the ground, which is similar to the change in pressure made in the action of getting up. Furthermore, the misclassification between Downstairs and some Walking activities can be explained by the nature of the dataset, which included subjects collecting data in the wild

Table 2 Cumulative confusion matrix of the DeepHAR against the testing dataset using a stratified tenfold cross-validation strategy

Predicted → Reference ↓	Cross trainer	Downstairs	Fast walking	Free walking	Normal walking	Sit to stand	Sitting	Slow walking	Spinning bike	Standing	Stretch	Upstairs
Cross trainer	**724**	0	0	0	0	0	0	0	3	0	1	1
Downstairs	0	**311**	0	4	1	12	0	1	0	0	1	24
Fast walking	0	7	**1849**	0	6	2	0	0	0	0	0	3
Free walking	0	0	0	**121**	0	0	0	1	0	0	0	8
Normal walking	0	4	9	2	**1809**	0	0	7	0	1	0	4
Sit to stand	0	0	0	0	0	**293**	0	0	0	2	0	1
Sitting	0	0	0	0	0	0	**248**	0	0	0	0	0
Slow walking	0	3	2	2	2	1	0	**1856**	0	0	0	3
Spinning bike	0	0	0	0	0	0	0	0	**2199**	0	0	0
Standing	0	0	0	0	0	2	0	0	0	**438**	0	1
Stretch	0	3	0	1	0	0	0	0	0	0	**204**	1
Upstairs	0	19	0	2	0	4	0	1	1	0	0	**406**

Bold values refers the principal diagonal of the confusion matrix

and recording session of Downstairs activities by ascending several stairs while walking on landings between floors. For Free Walking activities, there was a high rate of misclassification with other Walking activities. Although the Free Walking activities have been collected by the user with the freedom to walk in any direction without constraints, they necessarily combine the different walking speeds by creating an overlap between them. However, different walking activities have been included in the study for scenarios where the solution wants to be used for the rehabilitation of patients, where ambulation capabilities have to be analysed. Overall, fitness activities were recognised at a higher rate than ambulation activities; however, treating walking activities as the only activity may improve the prediction.

To evaluate the impact of data pre-processing on the performance of the proposed solution, a comparison was made between its performance with and without pre-processing. Additionally, to validate the proposed architecture, it was compared against a multi-layer perceptron (MLP), which is considered a basic feed-forward neural network composed of only input, output, and one hidden layer. As shown in Fig. 6, the proposed solution's core architecture outperforms the MLP solution, with an Accuracy of 96.89% compared to 91.99% for the MLP. However, by incorporating data pre-processing techniques, an even greater improvement in performance can be observed, with

the Accuracy reaching 98.56%. This comparison demonstrates that the use of pre-processing techniques not only improves the performance of the proposed solution but also enables a simpler architecture such as the feed-forward network to compete with state-of-the-art solutions that utilise more complex architectures.

4.1 Comparison with state-of-the-art solutions

Considering the advances achieved in the literature, four studies [24, 25, 27, 28] have been selected, which provided enough information to be retrained on the available dataset,
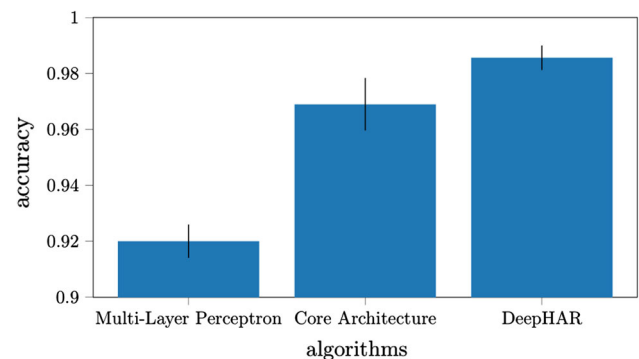

Fig. 6 Comparison of the precision performance of feed-forward neural networks. By core architecture, it is meant the neural network architecture proposed in this study without any optimisation

for comparison with the proposed DeepHAR solution. Deep learning and shallow machine learning were both covered in the studies that were chosen. Studies on machine learning were chosen because they contributed to the development of popular models like random forests [24] and SVM [25], while studies on deep learning included CNN networks, which are currently the most popular despite their complexity. Particularly in this latter instance, the two CNNs differ in how the data are handled since in one research, the data are processed by a different network for each modality before being combined [27], but in the other, they are processed simultaneously [28].

These algorithms have been trained using our dataset, however, remaining invariant with the number of sensors included, but applying the settings defined in the related papers. The settings involved for each experiment are reported in Table 3.

The results obtained from the comparative analysis are presented in Table 4. The solution proposed in this paper outperformed the other solutions analysed. Overall, the solutions based on deep learning outperformed those based on shallow machine learning, even if in the latter an engineering of the features has been employed, highlighting the effectiveness of using deep learning for the analysis of raw sensor data. The solution proposed by Wang et al. [28] exhibited performance that is comparable to that of the proposed solution, but with a higher standard deviation, indicating that its results were more heavily dependent on the samples included in the test set during cross-validation. By contrast, the proposed solution's use of a loss function that penalises the majority class during training allows it to handle the imbalanced dataset. Moreover, the importance of data pre-processing can be further identified by this comparison, because, under equal settings conditions, such as the work proposed by Pham et al. [27], in which an identical time window was used, our solution manages to obtain better performance even if the neural network used is simpler.

4.2 Study limitations identified

While the results are promising for real-life scenario applications, the following limitations have been identified for further work. The proposed architecture, comprised of smart insoles, mobile application, and cloud server, is in a prototype state and is currently focused on data collection and data storage. The use of cloud storage made it possible to collect data from study participants in an agile way and to periodically update the data with which the model was trained, obtaining better performance. Alternatives solutions to the cloud, such as the embedding of the activity recognition algorithm directly on the edge device (e.g. the smartphone) can be adopted. However, it has a number of drawbacks, including the need for larger memory capacity and increases computational costs on the edge device, a decrease in algorithm performance, and the inability to update the model as new data are gathered. Furthermore, the connection between smart insoles and the smartphone has been provided by Bluetooth Low Energy; nevertheless, in a future study, additional transmission technologies will be explored such as Wi-Fi and ZigBee, which could provide additional benefits in indoor environments.

The activities involved in this study comprised ambulation and fitness activities. Although their classification has been adequately achieved by the proposed solution, considering the walking activities at different speeds has affected the final performance of the solution; hence, combining them into a single activity can improve the performance. Enhancing the set of activities with further fitness activities, such as running or jogging, and daily living activities can provide a way to develop a thorough monitoring system for the subject's daily life. Moreover, transitioning between activities and interleaving between them have not been entirely addressed, and while the overlapping windows have lessened these two concerns, they still require additional examination. The dataset used is characterised by data collected from only five participants, so there is the risk of misclassification when using this solution with data obtained from people who have no resemblance to those analysed. The next stage, therefore,

Table 3 Settings used in the selected studies for the state-of-the-art performance comparison

Study	Algorithm	Sampling frequency (Hz)	Window size
Presented work	Feed-forward NN	200 downsampled to 20 by averaging 10 contiguous samples	2 s, 50% overlap
Pham et al. [27]	Multi-input CNN	20	2 s, 50% overlap
Wang et al. [28]	1D-CNN	50	6 s
De Pinho et al. [24]	RF	10	0.3 s
Sazonov et al. [25]	SVM	400 downsampled to 25 by averaging 16 contiguous samples	2 s

Table 4 Results of the state-of-the-art performance comparison. For all experiments, a stratified tenfold cross-validation was used

Study	Accuracy (%)	F1-Score (%)	AUC score (%)
Presented work	**98**.56 (±0.44)	**98**.57 (±0.46)	**99**.25 (±0.23)
Pham et al. [27]	97.47 (±0.48)	97.41 (±0.55)	96.37 (±0.89)
Wang et al. [28]	98.34 (±1.18)	98.33 (±1.22)	97.20 (±1.82)
De Pinho et al. [24]	96.71 (±0.40)	96.70 (±0.43)	98.09 (±0.25)
Sazonov et al. [25]	94.98 (±0.48)	94.62 (±0.51)	97.26 (±0.27)

Bold values refers the best-performing model

will be to collect more data from heterogeneous people, including various ages and ethnicities, in order to promote subject-independent learning and determine whether there is a relation between participants' characteristics and the way the activities are carried out. Additionally, the expansion of the dataset may favour the implementation of a leave-one-subject-out cross-validation technique to assess the model's performance. This strategy enables testing the solution on data from a subject that was not utilised during training, demonstrating the solution's generalisation. Since the suggested algorithm is based on data collected by smart insoles worn on both legs, the results may be unreliable if the user of the system is unable to wear both, such as in the case of a lower-limb amputee. Therefore, additional analysis will be performed to categorise those activities involving only one leg.

5 Conclusion

In this paper, a smart insole-based human activity recognition solution for ambulation and fitness activities has been presented. The smart insole, comprised of pressure and inertial sensors, has been used as the only device to make the solution non-invasive for the user. Without using any heuristic feature extraction techniques, a deep feed-forward neural network method has been proposed for processing directly the raw data and forecasting the activities. The proposed solution achieved an Accuracy and F1-Score of 98.56% and 98.57%, respectively. The Sitting activities obtained the highest degree of recognition, with 100% Precision and Sensitivity, closely followed by the Spinning Bike class, which achieved a Precision and Sensitivity of 99.82% and 100%, respectively. Overall, fitness activities were recognised at a higher rate than ambulation activities, which were affected by multiple misclassifications due to the stairs activities and overlap between the various walking activities. Although there are some issues in differentiating Downstairs from Upstairs activities, the model has a high generalisation rate between classes as demonstrated by the overall AUC value which is 99.25%. Even though the integration of both free walking and

walking at various speeds led to overlaps that had an impact on the classifier's performance, it should be noted that these activities are fundamental for rehabilitation monitoring because they allow for the estimation of a patient's degree of ambulation. The deep feed-forward neural network proposed in this study has been enhanced by data pre-processing techniques, including data interpolation for handling missing data, data segmentation of 2 s with overlapping of 50%, and signal down-sampling by use of the averaging technique for noise reduction. Moreover, to handle the imbalanced dataset, a cost-sensitive re-weighting approach has been involved to update the loss function of the proposed model, penalising the majority classes by using small weights and favouring the minority classes by greater weights. To evaluate the effect of pre-processing on the performance of the proposed deep learning solution, a comparative analysis has been carried out with and without pre-processing. The solution has been compared further with a multi-layer perceptron (MLP) as a basic feed-forward neural network. The results showed that the proposed solution's core architecture outperformed the MLP, with an Accuracy of 96.89% compared to 91.99% for the MLP. However, by incorporating pre-processing techniques, the accuracy improved even further to 98.56%. Furthermore, to better ascertain the capabilities of the proposed solution, the results were compared with state-of-the-art solutions trained with the same dataset, outperforming them. This comparison demonstrates that using pre-processing not only improves performance but also allows for a simpler architecture to compete with more advanced solutions, making the solution feasible for health monitoring and/or rehabilitation applications while reducing computational costs.

6 Future work

One of the key issues encountered in this study has been the lack of individuals available to gather the data, which made it difficult to examine how demographics and other personal traits might affect the performance of the model. Although in the literature multiple analysis has been

carried out in determining the effects of genetics, cultural practices, and demography on the gait [56, 57], there is no analysis in determining the impact that those factors have during carrying out different activities. Furthermore, considering that even the same individual can perform the same activity in different ways and that the proposed solution is a data-driven solution, that relies mainly on the data analysed, the first step in future research will be to broaden the participant cohort, accounting for greater differentiation and more age groups and diverse cultures. With the aim of providing a system that can be used on a daily basis, future research will include additional activities, such as running or jogging as well as daily activities. Having multiple sensors available within the smart insoles results in high energy expenditure, therefore, a future study will focus on analysing the importance that each sensor has on the classification and a minimum configuration will be sought to reduce such consumption. Furthermore, given the misclassifications in stair-related activities, the impact of introducing a barometer into the proposed system for evaluating altitude changes will be analysed.

Data availability The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval This research received ethics approval from the CEBE Faculty Research Ethics Committee, Ulster University (reference number: CEBE_RE_20.11).

References

1. Saho K, Hayashi S, Tsuyama M, Meng L, Masugi M (2022) Machine learning-based classification of human behaviors and falls in restroom via dual doppler radar measurements. Sensors 22:1721. https://doi.org/10.3390/S22051721
2. Marques JB, Mc Auliffe S, Thompson A, Sideris V, Santiago P, Read PJ (2022) The use of wearable technology as an assessment tool to identify between-limb differences during functional tasks following acl reconstruction. A scoping review. Phys Ther Sport 55:1–11. https://doi.org/10.1016/j.ptsp.2022.01.004
3. Zhang P, Zhang J (2022) Deep learning analysis based on multi-sensor fusion data for hemiplegia rehabilitation training system for stoke patients. Robotica 40(3):780–797. https://doi.org/10.1017/S0263574721000801
4. Elshafei M, Costa DE, Shihab E (2022) Toward the personalization of biceps fatigue detection model for gym activity: an approach to utilize wearables' data from the crowd. Sensors. https://doi.org/10.3390/s22041454
5. Li S, Zheng P, Fan J, Wang L (2022) Toward proactive human-robot collaborative assembly: a multimodal transfer-learning-enabled action prediction approach. IEEE Trans Industr Electron 69(8):8579–8588. https://doi.org/10.1109/TIE.2021.3105977
6. Xiao W, Liu H, Ma Z, Chen W (2022) Attention-based deep neural network for driver behavior recognition. Future Gener Comput Syst 132:152–161. https://doi.org/10.1016/j.future.2022.02.007
7. Saleem G, Bajwa UI, Raza RH (2022) Toward human activity recognition: a survey. Neural Comput Appl 2022:1–38. https://doi.org/10.1007/S00521-022-07937-4
8. Kulsoom F, Narejo S, Mehmood Z, Chaudhry HN, Butt A, Bashir AK (2022) A review of machine learning-based human activity recognition for diverse applications. Neural Comput Appl 34:21–34, 18289–18324. https://doi.org/10.1007/S00521-022-07665-9
9. Kumar KV, Harikiran J (2022) Privacy preserving human activity recognition framework using an optimized prediction algorithm. IAES Int J Artif Intell 11(1):254–264. https://doi.org/10.11591/ijai.v11.i1.pp254-264
10. Dang LM, Min K, Wang H, Piran MJ, Lee CH, Moon H (2020) Sensor-based and vision-based human activity recognition: a comprehensive survey. Pattern Recogn 108:107561. https://doi.org/10.1016/j.patcog.2020.107561
11. Qiu S, Zhao H, Jiang N, Wang Z, Liu L, An Y, Zhao H, Miao X, Liu R, Fortino G (2022) Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. Inf Fusion 80:241–265. https://doi.org/10.1016/j.inffus.2021.11.006
12. Chen D, Cai Y, Qian X, Ansari R, Xu W, Chu K-C, Huang M-C (2019) Bring gait lab to everyday life: gait analysis in terms of activities of daily living. IEEE Internet Things J 7(2):1298–1312. https://doi.org/10.1109/JIOT.2019.2954387
13. Aznar-Gimeno R, Labata-Lezaun G, Adell-Lamora A, Abadía-Gallego D, del-Hoyo-Alonso R, González-Muñoz C (2021) Deep learning for walking behaviour detection in elderly people using smart footwear. Entropy 23(6):777. https://doi.org/10.3390/e23060777
14. Kalimuthu S, Perumal T, Yaakob R, Marlisah E, Babangida L (2021) Human activity recognition based on smart home environment and their applications, challenges. In: 2021 international conference on advance computing and innovative technologies in engineering (ICACITE). IEEE, pp 815–819. https://doi.org/10.1109/ICACITE51222.2021.9404753
15. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal

Mach Intell 35(8):1798–1828. https://doi.org/10.48550/arXiv.1206.5538

16. Gupta P, McClatchey R, Caleb-Solly P (2020) Tracking changes in user activity from unlabelled smart home sensor data using unsupervised learning methods. Neural Comput Appl 32:12351–12362. https://doi.org/10.1007/S00521-020-04737-6

17. Noor MHM (2021) Feature learning using convolutional denoising autoencoder for activity recognition. Neural Comput Appl 33:10909–10922. https://doi.org/10.1007/S00521-020-05638-4

18. Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y (2021) Deep learning for sensor-based human activity recognition: overview, challenges, and opportunities. ACM Comput Surv. https://doi.org/10.1145/3447744

19. Moticon ReGo AG (2023) OpenGo sensor insoles. https://moticon.com/opengo/sensor-insoles. Accessed 26 Jan 2023

20. IEE Luxemburg SA (2023) Smart footwear. https://iee-sensing.com/health-tech/medical/smart-footwear-sensing-solutions/. Accessed 26 Jan 2023

21. Salted Ltd. (2023) Neurogait insoles. https://www.salted.ltd/eng/main/index.html. Accessed 26 Jan 2023

22. Ngueleu AM, Blanchette AK, Maltais D, Moffet H, McFadyen BJ, Bouyer L, Batcho CS (2019) Validity of instrumented insoles for step counting, posture and activity recognition: a systematic review. Sensors 19:2438. https://doi.org/10.3390/S19112438

23. Moufawad el Achkar C, Lenoble-Hoskovec C, Paraschiv-Ionescu A, Major K, Büla C, Aminian K (2016) Instrumented shoes for activity classification in the elderly. Gait Posture 44:12–17. https://doi.org/10.1016/j.gaitpost.2015.10.016

24. De Pinho André R, Diniz PHFS, Fuks H (2017) Bottom-up investigation: human activity recognition based on feet movement and posture information. In: Proceedings of the 4th international workshop on sensor-based activity recognition and interaction. iWOAR '17. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3134230.3134240

25. Sazonov E, Hegde N, Browning RC, Melanson EL, Sazonova NA (2015) Posture and activity recognition and energy expenditure estimation in a wearable platform. IEEE J Biomed Health Inform 19(4):1339–1346. https://doi.org/10.1109/JBHI.2015.2432454

26. D'Arco L, Wang H, Zheng H (2022) Assessing impact of sensors and feature selection in smart-insole-based human activity recognition. Methods Protocols. https://doi.org/10.3390/mps5030045

27. Pham C, Diep NN, Phuong TM (2017) e-shoes: smart shoes for unobtrusive human activity recognition. In: 2017 9th international conference on knowledge and systems engineering (KSE). IEEE, pp. 269–274. https://doi.org/10.1109/KSE.2017.8119470

28. Wang L, Peng M, Zhou QF (2019) Fall detection based on convolutional neural networks using smart insole. In: 2019 5th international conference on control, automation and robotics (ICCAR). IEEE, pp. 593–598. https://doi.org/10.1109/ICCAR.2019.8813332

29. Paydarfar AJ, Prado A, Agrawal SK (2020) Human activity recognition using recurrent neural network classifiers on raw signals from insole piezoresistors. In: 2020 8th IEEE RAS/EMBS international conference for biomedical robotics and biomechatronics (BioRob). IEEE, pp. 916–921. https://doi.org/10.1109/BioRob49111.2020.9224311

30. D'Arco L, Wang H, Zheng H (2021) Artificial neural network for human activity recognition by use of smart insoles. In: Proceedings of the 7th collaborative European research conference (CERC 2021), Cork, Ireland

31. Banan A, Nasiri A, Taheri-Garavand A (2020) Deep learning-based appearance features extraction for automated carp species identification. Aquacult Eng 89:102053. https://doi.org/10.1016/j.aquaeng.2020.102053

32. Afan HA, Osman AIA, Essam Y, Ahmed AN, Huang YF, Kisi O, Sherif M, Sefelnasr A, Chau K-W, El-Shafie A (2021) Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques. Eng Appl Comput Fluid Mech 15(1):1420–1439. https://doi.org/10.1080/19942060.2021.1974093

33. McCalmont G, Morrow P, Zheng H, Samara A, Yasaei S, Wang H, McClean S (2018) ezigait: toward an ai gait analysis and sssistant system. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM). pp. 2280–2286. https://doi.org/10.1109/BIBM.2018.8621176

34. Lara OD, Labrador MA (2012) A survey on human activity recognition using wearable sensors. IEEE Commun Surveys Tutor 15(3):1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192

35. Hegde N, Bries M, Swibas T, Melanson E, Sazonov E (2017) Automatic recognition of activities of daily living utilizing insole-based and wrist-worn wearable sensors. IEEE J Biomed Health Inf 22(4):979–988. https://doi.org/10.1109/JBHI.2017.2734803

36. Pham C, Nguyen-Thai S, Tran-Quang H, Tran S, Vu H, Tran T-H, Le T-L (2020) Senscapsnet: deep neural network for non-obtrusive sensing based human activity recognition. IEEE Access 8:86934–86946. https://doi.org/10.1109/ACCESS.2020.2991731

37. Tiu ESK, Huang YF, Ng JL, AlDahoul N, Ahmed AN, Elshafie A (2022) An evaluation of various data pre-processing techniques with machine learning models for water level prediction. Nat Hazards 110:121–153. https://doi.org/10.1007/S11069-021-04939-8

38. Wang H, Li S, Song L, Cui L, Wang P (2020) An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network. IEEE Trans Instrum Meas 69:2648–2657. https://doi.org/10.1109/TIM.2019.2928346

39. Pitaloka DA, Wulandari A, Basaruddin T, Liliana DY (2017) Enhancing cnn with preprocessing stage in automatic emotion recognition. Procedia Comput Sci 116:523–529. https://doi.org/10.1016/J.PROCS.2017.10.038

40. Salgado CM, Azevedo C, Proença H, Vieira SM (2016) Secondary Analysis of Electronic Health Records. Springer, New York, pp 143–162. https://doi.org/10.1007/978-3-319-43742-2_13

41. Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O (2021) A survey on missing data in machine learning. J Big Data 8(1):1–37. https://doi.org/10.1186/s40537-021-00516-9

42. Lepot M, Aubin J-B, Clemens FHLR (2017) Interpolation in time series: an introductive overview of existing methods, their performance criteria and uncertainty assessment. Water. https://doi.org/10.3390/w9100796

43. Hegde N, Zhang T, Uswatte G, Taub E, Barman J, McKay S, Taylor A, Morris DM, Griffin A, Sazonov ES (2017) The pediatric smartshoe: wearable sensor system for ambulatory monitoring of physical activity and gait. IEEE Trans Neural Syst Rehabil Eng 26(2):477–486. https://doi.org/10.1109/TNSRE.2017.2786269

44. Merry KJ, Macdonald E, MacPherson M, Aziz O, Park E, Ryan M, Sparrey CJ (2021) Classifying sitting, standing, and walking using plantar force data. Med Biol Eng Comput 59(1):257–270. https://doi.org/10.1007/s11517-020-02297-4

45. Quigley B, Donnelly M, Moore G, Galway L (2018) A comparative analysis of windowing approaches in dense sensing environments. Proceedings. https://doi.org/10.3390/proceedings2191245

46. Banos O, Galvez J-M, Damas M, Pomares H, Rojas I (2014) Window size impact in human activity recognition. Sensors 14(4):6474–6499. https://doi.org/10.3390/s140406474

47. Putra IPES, Vesilo R (2017) Window-size impact on detection rate of wearable-sensor-based fall detection using supervised machine learning. In: 2017 IEEE life sciences conference (LSC). pp 21–26. https://doi.org/10.1109/LSC.2017.8268134

48. Lee KS, Chae S, Park HS (2019) Optimal time-window derivation for human-activity recognition based on convolutional neural networks of repeated rehabilitation motions. In: IEEE international conference on rehabilitation robotics: [proceedings]. pp 583–586. https://doi.org/10.1109/ICORR.2019.8779475

49. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Teh YW, Titterington M (eds) Proceedings of the thirteenth international conference on artificial intelligence and statistics. Proceedings of machine learning research, vol 9. PMLR, Chia Laguna Resort, Sardinia pp 249–256

50. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958

51. Buda M, Maki A, Mazurowski MA (2018) A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 106:249–259. https://doi.org/10.1016/j.neunet.2018.07.011

52. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) Smote: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357. https://doi.org/10.1613/jair.953

53. Drummond C, Holte RC, et al (2003) C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In: Workshop on learning from imbalanced datasets II, vol 11. Citeseer, pp 1–8

54. Huang C, Li Y, Loy CC, Tang X (2016) Learning deep representation for imbalanced classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5375–5384. https://doi.org/10.1109/CVPR.2016.580

55. Khan SH, Hayat M, Bennamoun M, Sohel FA, Togneri R (2017) Cost-sensitive learning of deep feature representations from imbalanced data. IEEE Trans Neural Netw Learn Syst 29(8):3573–3587. https://doi.org/10.1109/TNNLS.2017.2732482

56. Chen W-L, O'Connor JJ, Radin EL (2003) A comparison of the gaits of Chinese and Caucasian women with particular reference to their heelstrike transients. Clin Biomech 18(3):207–213. https://doi.org/10.1016/S0268-0033(02)00187-0

57. Boulifard DA, Ayers E, Verghese J (2019) Home-based gait speed assessment: normative data and racial/ethnic correlates among older adults. J Am Med Dir Assoc 20(10):1224–1229. https://doi.org/10.1016/j.jamda.2019.06.002