This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

A Human Activity Recognition Method Based on Lightweight Feature Extraction Combined with Pruned and Quantized CNN for Wearable Device

MYUNG-KYU YI, WAI-KONG LEE, *Member, IEEE*, and SEONG OUN HWANG, *Senior Member, IEEE*

***Abstract*—Human Activity Recognition (HAR) is becoming an essential part of human life care. Existing HAR methods are usually developed using a two-level approach, wherein a first-level Machine Learning (ML) classifier is employed to distinguish the static and dynamic activities, followed by a second-level classifier to identify the specific activity. These approaches are not suitable for wearable devices, due to the high computational and memory consumption. Our rigorous analysis of various HAR datasets opens up a new possibility that static or dynamic activities can be discriminated against through a simple statistical technique. Therefore, we propose to utilize a statistical feature extraction technique to replace the first-level ML classifier, thus achieving more lightweight computation. Next, we employ Random Forest (RF) and Convolutional Neural Networks (CNN) to classify the specific activities, achieving higher accuracy compared to the state-of-the-art results. We further reduce the computation and memory consumption of the above combined approach by applying pruning and quantizing techniques to CNN (PQ-CNN). Experimental results show the proposed lightweight HAR method achieved an F1 score of 0.9417 and 0.9438 for unbalanced and balanced datasets, respectively. On top of lightweight and accuracy, the proposed HAR method is practical for wearable devices by using a single accelerometer.**

***Index Terms*—human activity recognition, convolutional neural network, wearable sensors, deep learning.**

I. Introduction

The demand of consumers to improve the quality of life is steadily increasing. This inspired many research activities in the healthcare and fitness area, centering around human activity monitoring. Wearable IoT systems that can recognize human physical activity and behavior are getting more popular nowadays. Human Activity Recognition (HAR) is the essential technology for identifying and interpreting human physical activity and behavior in these wearable IoT systems. It first collects data from wearable or optical sensors, then recognizes and analyzes human activities such as sleep patterns or health conditions. By incorporating the collected HAR data with other medical information like blood pressure, heart rate, and blood sugar level, a doctor can analyze and interpret the data for diagnosis, prescribing medicine, or recommending treatment. Based on the types of collected data, there are two main approaches to HAR systems: sensor-based activity recognition and vision-based activity recognition. To identify human activities, a sensor-based HAR collects and analyzes data from wearable sensors such as accelerometers or gyroscopes that are attached to the body. Vision-based HAR extracts and analyzes features from images or videos obtained from optical sensors without wearable sensors. However, it is very difficult to identify human behavior exactly through videos taken by various cameras due to variations in human shapes and motions, cluttered backgrounds, and viewpoint variations. Therefore, the sensor-based HAR is becoming more popular and widely used in various domains.

Despite huge advances being achieved by sensor-based HAR methods [1], the existing sensor-based HAR methods have mainly focused on maximizing the accuracy of the HAR methods. The use of conventional Deep Learning (DL) models with more filters, layers, and sensors may improve the accuracy of the HAR, but they increase power consumption and computational complexity proportionally [2]. Since the bigger neural networks get, the more complex they are. Therefore, more computation power is required to train data for the DL model effectively.

Some researchers have sought the lightweight HAR approach, which can be suitable to be executed on edge devices [2], [3]. However, these lightweight HAR studies which are focused on smartphones have limitations. Referring to Table I, smartphones have more resources to perform many tasks (e.g., incoming phone calls, connecting to Wi-Fi networks, managing software, etc). However, wearable devices limited in processing power, energy, and storage space can only handle simple tasks compared to smartphones. Therefore, existing smartphone-based solutions that rely on complicated computations using a deep CNN and multiple sensors may not be suitable for wearable devices. This motivates further research to be carried out to fulfill such demands: the HAR method should be accurate and require less computation with a smaller amount of memory to be fit for the constraints of wearable devices.

In the future, wearable devices are expected to become more ubiquitous, smaller, and cheaper. Currently, most wearable devices are prominent such as smartwatches, smart bands, smart glasses, and smart shoes, but they will develop into tiny wearable devices such as smart rings, e-skin, and smart tattoos. Tiny wearable devices could provide biometric data

This work was supported by the High-Potential Individuals Global Training Program of the Institute of Information communications Technology Planning & Evaluation (IITP) grant (2021-0-00540, 40%), and National Research Foundation of Korea (NRF) grants (2019H1D3A1A01102607, 30% and 2020R1A2B5B01002145, 30%) funded by the Korean Government through Ministry of Science and ICT (MSIT).

Corresponding authors: Seong Oun Hwang (sohwang@gachon.ac.kr)

MYUNG-KYU YI, WAI-KONG LEE, and SEONG OUN HWANG are with the Department of Computer Engineering, Gachon University, Seongnam 13120, South Korea

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

TABLE I
OVERVIEW OF RESOURCES AVAILABLE IN SMARTPHONES AND WEARABLE DEVICES

	CPU Clock	RAM	Sensors	Battery
Smartphones	Multi-core with few Giga Hz (Mongoose M4 [4])	Giga bytes	Complex sensors (e.g., camera)	Recharge every 3-4 days
Wearable Devices	Single-core with less than 500 Mega Hz (e.g., STM32H7 [5])	1 mega bytes	Limited low-cost sensors (e.g., accelerometer)	Recharge every month

and activity levels from a HAR chip inside a user's ring or shirt button. They can even provide alarms to warn wrist fatigue when someone is using the keyboards for a long time. A lightweight Artificial Intelligence (AI) model for tiny wearable devices can greatly reduce the burden on the wearer and improve the user experience, which is the future direction of the HAR system development. To implement practical HAR on wearable devices, the HAR system must not only collect data continuously but also be able to respond quickly in real time. Moreover, wearable devices are used to perform monitoring tasks (including HAR) for a relatively long period. To satisfy the requirements above, AI models should be embedded in the wearable device itself, not in smartphones or server platforms. Hence, wearable HAR must employ a simple DL architecture to reduce memory and computational complexity so that it can fit into resource-constrained wearable devices. Consequently, it is necessary to design an energy-efficient AI model while maintaining relatively high accuracy considering the limitations of wearable devices.

Some earlier HAR techniques developed based on ML techniques such as Decision Tree (DT), Support Vector Machine (SVM), and Random Forests (RF) have been used to identify the activity patterns of users given the data captured from wearable sensors [1], [6], [7]. However, ML techniques require handcrafted feature extraction methods to extract features from the dataset. This can be a serious limitation as domain knowledge is required to develop these feature extraction methods. In contrast, DL techniques are capable of learning features automatically. Therefore, sensor-based HAR explored the task of activity recognition using single-level techniques, such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) [1], [8], [9].

Several works have shown efforts to achieve high-level accuracy in HAR using smartphones and powerful multiple sensors, such as accelerometers, gyroscopes, and magnetometers [10]–[12]. However, the use of multiple sensors for the HAR system requires more power consumption. For example, the use of a gyroscope can help improve the recognition rates of HAR, but that consumes six times the power of an accelerometer [13]. Multiple sensors are more expensive than a single sensor. Another way to increase accuracy in HAR systems is to use additional layers in the DL model. Teng et al. [14] and Tang et al. [15] proposed local loss-based CNN for HAR and the layer-wise training Lego CNN, respectively. Several methods such as hybrid, multi-branch, and ensemble HAR were proposed to improve the accuracy of HAR [16]–[27]. However, these methods greatly increase computational cost and complexity.

To obtain an optimized solution, many researchers have adopted a two-level approach - first separate static and dynamic activities using DT or SVM classifier, and then apply ML classifiers and DL to them [3], [28]–[30]. This is backed by the observation that 84% of the time in human daily life is spent on static activities, and only 16% of the time is spent on dynamic activities [29], [31]. Based on this approach, the two-level HAR methods can reduce power consumption and computational complexity. However, such an approach poses a big challenge because it requires an additional classifier such as SVM or DT at the first level, which can be a costly overhead for embedded systems. This motivates us to investigate the possibility of replacing the first-level ML classifier with a more lightweight approach, which can be beneficial for the deployment of HAR on wearable devices.

In this paper, we analyzed that the human body consists of several body segments such as the head, arms, legs, and torso. The directions and positions of the body segments do not change significantly in static activities such as sitting, standing and lying. However, the situation is completely the opposite in dynamic activities such as walking, running, and jumping. Therefore, when the accelerometer is attached to the body segments, it is more likely to generate stable readings for static activities and time-varying readings for dynamic activities. Due to this reason, the static activities can be identified easily by only observing the positions and directions of the accelerometer readings. This offers a possibility that static activities can be easily discriminated by simple technique like standard variance. This idea can also be generically applicable to any HAR datasets because the nature of accelerometer readings generated from static (stable) and dynamic (fluctuating significantly) activities is the same. To validate our insight, we rigorously analyzed the characteristics of various open datasets such as UCI-HAR [32], WISDM [33], PAMAP [34], DSADS [35] and UniMiB SHAR [36] first, and finally verified that static and dynamic activities can be distinguished based on the standard deviation values of the accelerometer.

Based on this insightful observation, we proposed a new two-level HAR method in this paper. The commonly used ML classifier at the first level is replaced by a statistical feature extraction technique. We also proposed to combine RF with pruned and quantized CNN (PQ-CNN) utilizing data obtained from only one accelerometer. Our contributions are summarized as follows.

1) Existing two-level HAR methods utilize machine learning classifiers in the first stage to classify static and dynamic activities, which requires long training and pro-

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

cessing time. Through analysis of various representative datasets, we discovered that this classification task can be completed through a simple statistical technique. In this paper, we propose a lightweight feature extraction technique that can complete the classification of static and dynamic activities within a few seconds, which is 5 times and 32.95 times faster than SVM and DT classifiers, respectively. This technique only depends on simple statistical information without any training required. In addition, the proposed technique also maintains a higher recognition rate, on average, compared to SVM or DT classifiers.

2) To the best of our knowledge, this paper is the first to combine an RF classifier and a PQ-CNN classifier for HAR. The existing two-level HAR approaches fail to achieve high accuracy because they used the lightweight, less accurate ML classifier like DT in classifying static activities. In contrast, we proposed to use RF, because it demonstrates the highest accuracy among all the commonly used ML models in HAR for recognizing static activities. On top of that, we utilize CNN for classifying complex dynamic activities, which shows very high accuracy. By combining RF and CNN, the proposed HAR method achieves higher accuracy compared to state-of-the-art HAR methods.

3) To further reduce the memory and computational overhead, we proposed to prune and quantize the CNN. This combination (RF + PQ-CNN) is optimal because it has relatively low memory and computational overhead compared to existing hybrid HAR methods. Besides being lightweight, the proposed HAR method only uses one accelerometer, which makes the solution wearable-ready. It can be directly implemented on low-power wearable hardware [5] that comes with extremely limited resources (e.g., the number of sensors, power, memory) Compared to other solutions that rely on multiple sensors and consume large memory space, our method is more practical for wearable devices that need to perform continuous monitoring tasks for a long period.

The rest of the paper is organized as follows. Section II reviews related work on HAR. Section III describes the details of the proposed HAR method. Section IV presents and analyzes the experimental results and gives discussions. Section V concludes the paper.

II. Related Work

In order to improve the accuracy of HAR, state-of-the-art works in HAR applications [3], [28]–[30] have focused on the two-level HAR method to reduce power consumption. Cho and Yoon [28] presented a divide and conquer-based one-dimensional CNN (1D CNN) for HAR by employing test data sharpening. They used a DT to classify dynamic and static activity at the first stage. Then, the 1D CNN model is used to classify more detailed activity at the second stage. Bhat et al. [29] presented the two-level hardware accelerator to integrate a wide range of human activity recognition, where activities are classified as static or dynamic at the first stage by an SVM classifier, and then, a DT or a Deep Neural Network (DNN) classifier is used to identify further activities at the second stage. Coelho et al. [3] presented a lightweight framework for wearable devices with constrained computing resources. They used a two-level classifier to identify human activities. During the first phase, activities are classified as static or dynamic with a DT. Then, static and dynamic activity classification is executed by the DT or a CNN, respectively. Huang et al. [30] presented a two-stage end-to-end CNN with a data augmentation method. Using two cascaded CNNs, walking upstairs and downstairs are mainly classified by the first level CNN, with other activities classified at the second stage. The two-level HAR can reduce power consumption by up to two times, compared to the traditional HAR systems [3], [29]. But the disadvantage with the two-level HAR method is the cost for additional classifiers, such as an SVM or a DT, at the first level. The SVM is a widely used method that can treat problems involving small samples, but it still consumes significant memory and computation . Moreover, an SVM is not appropriate for classifying large datasets [37].

Patil et al. [19] and Semwal et al. [20] presented the performance of ELM (Extreme Learning Machine) neural network based classifiers to identify activity. The ELM neural network can be used in the HAR method because of its very fast training capacity and good performance. But, the ELM is not able to train large dataset fast and efficiently due to its memory-residency, and high space and time complexity [38]. Therefore, the ELM based HAR method is not suitable for wearable devices considering the limited computing resources.

The most widely used classification methods in HAR systems are the ensemble and hybrid approaches. The ensemble learning classifier can be used to improve the performance of HAR [7], [21], [22]. Semwal et al. [7] introduced an ensemble learning approach for multiple-task human gait analysis and identification by applying different classification algorithms like SVM, KNN, ELM, and Principal Component Analysis (PCA). Semwal et al. [21] proposed an optimized hybrid deep learning for the classification of gait activities using four hybrid deep learning models namely, CNN-LSTM, CNN-GRU, LSTM-CNN, LSTM-GRU, and the ensemble of all models. Jain et al. [22] introduced a CNN-LSTM ensemble model for lower extremity activity by considering different walking styles on different terrains. The ensemble learning technique used to avoid over fitting by reducing the complexity of model, but it is expensive in terms of memory and computation [39].

The hybrid DL model can be used to improve feature extraction by combining the different DL models [6], [23]–[27]. Challa et al. [23] proposed a a multibranch model for HAR. In this proposed HAR model, a hybrid of convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) were used. Nidhi Dua et al. [24], [25] proposed a CNN-Gated Recurrent Unit (GRU) model for HAR by utilizing the robustness of CNNs in feature extraction and the advantages of GRU for time series data classification. Bijalwan et al. [26] proposed a hybrid deep learning approach comprising CNN, CNN-LSTM, and CNN-GRU models for upper limb exercise of post-stroke rehabilitation patients. Bijalwan et al. [27] proposed the design of

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

four deep learning-based HAR classification models, namely, Deep Neural Network (DNN), Bidirectional Long Short-Term Memory (BLSTM), CNN, ad CNN-LSTM for seven different activity classifications. Semwa et al. [6]. proposed optimized feature selection using bio-geography optimization technique for human walking activities recognition.

In summary, ensemble learning and hybrid DL model are effective by combining multiple learning methods to yield better performance. The most important aspect in designing HAR method for wearable devices is to keep high accuracy and a low memory/computational load at the same time. This seems to be difficult to achieve by using ensemble learning and hybrid DL model, because more computation is required to achieve high accuracy. Despite their high-prediction performance, the above mentioned HAR methods may not be directly used in wearable devices.

Zebin et al. [2] proposed a novel CNN to recognize five classes of static and dynamic activity. In [2], feature learning was carried out automatically through CNNs. Teng et al. [14] proposed a layer-wise CNN by employing local error signals to train a CNN layer by layer. They updated and optimized each hidden layer by implementing two different local learning signals for training a network with local loss. Tang et al. [15] proposed a CNN that can greatly reduce computational complexity while maintaining accuracy.

Huang et al. [8] proposed shallow CNNs with channel selectivity, which can be enacted as a dynamic pruning and rewiring process. To improve the efficiency of the CNN, each convolutional layer can select more important channels. Gao et al. [9] introduced a new multi-branch CNN using selective kernel convolution for HAR. To classify various human activities, the kernel convolution can select an appropriate receptive field size from among multiple branches. Ascioglu and Senol [10] presented a wearable wireless multi-sensor activity monitoring system. They showed that their Convolutional LSTM (ConvLSTM) model for HAR outperformed a CNN or LSTM model.

Xia et al. [40] proposed a novel HAR method by employing two LSTM layers with convolutional layers. To improve the accuracy of HAR, they used LSTM in combination with a CNN [40]. In addition, HAR methods can achieve better accuracy than a baseline CNN by using multiple layers and special filters with complicated algorithms. However, HAR is a difficult time series classification task. Various ML or DL techniques can be used to detect activities by reading and processing sensor data automatically, but some DL techniques are heavy in nature. With an increasing number of kernels or convolutional layers in the CNN model, the complexity and the number of parameters of the CNN model increased, leading to increased accuracy of the HAR model. Therefore, it is very difficult to implement HAR in a wearable device due to its limited computation capabilities, storage, and power.

III. Proposed Lightweight HAR Method

Previous works [3], [28]–[30] proposed two-level approach that first classify the human activities into stationary and dynamic using an ML classifier, then further classify them into specific activities. In this paper, we also follow the two-level approach, but to make it fit for wearable devices, we came up with a lightweight method of statistical feature extraction rather than using the ML classifier in the first level. The proposed statistical feature extraction technique is analyzed theoretically in Section III-B and applied to the proposed HAR method, which comprises of a RF classifier and a lightweight CNN. This section describes the proposed lightweight HAR method. The system architecture is introduced first, followed by the proposed statistical feature extraction, the RF classifier, and the pruned and quantized CNN model.

A. System Architecture

Fig. 1 shows the overall architecture of the proposed lightweight HAR method.

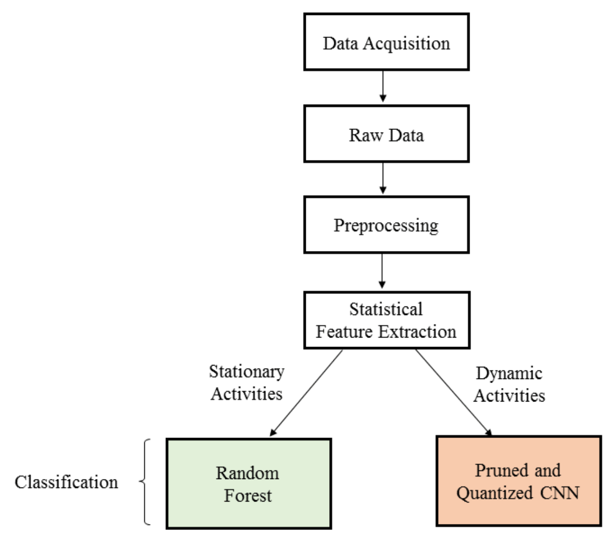

Fig. 1. Proposed lightweight HAR method

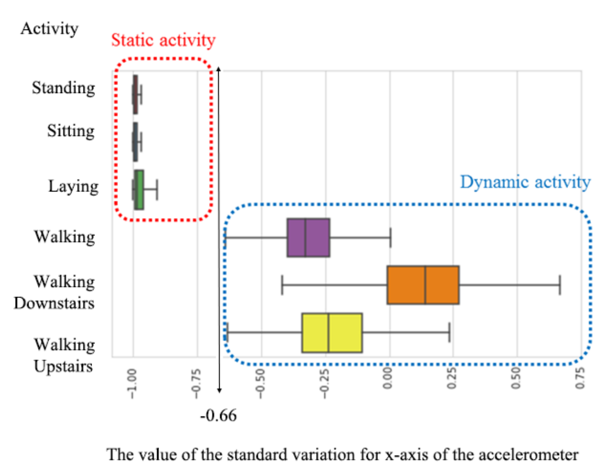

Fig. 2. Distribution of activities for $STDEV_X$ from the UCI dataset

Data collected from a tri-axial accelerometer are preprocessed. To extract valuable feature vectors for HAR effectively,

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

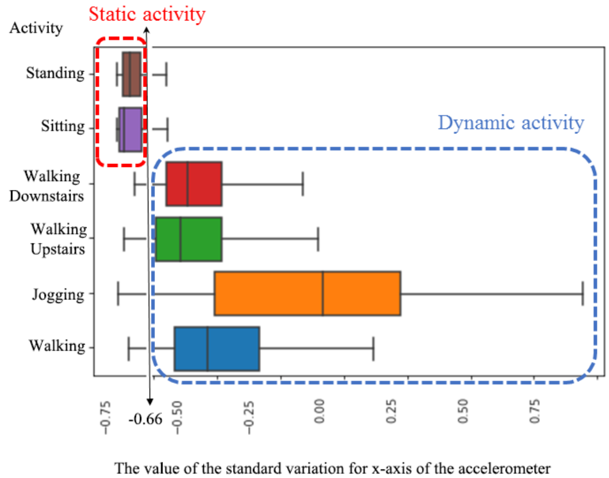

Fig. 3. Distribution of activities for $STDEV_X$ from the WISDM dataset

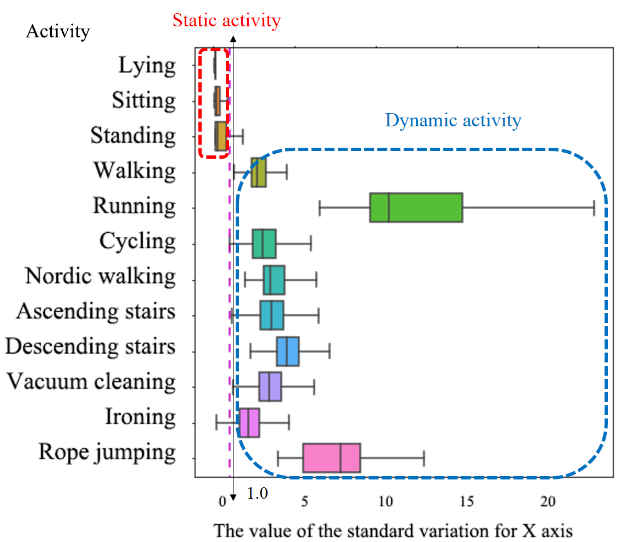

Fig. 4. Distribution of activities for $STDEV_X$ from the PAMAP dataset

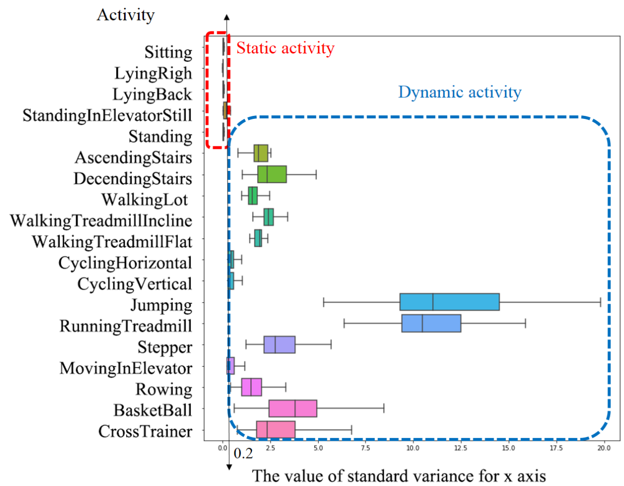

Fig. 5. Distribution of activities for $STDEV_X$ from the DSADS dataset

filter-based noise reduction approaches like Gaussian filter and adaptive filter in the frequency domain can be used [41], [42]. Then, data should be divided into smaller fragments called windows. Through this process, each window is considered an independent user activity. However, the window size directly affects activity recognition results. Each window can include information on multiple activities with a large window size. On the contrary, a small window size causes too many recognition tasks without achieving high recognition results. Sliding window is the most widely used technique for segmentation in HAR.

Various noise filters and segmentation techniques can provide a detailed recognition of each phase in a complex activity [43], [44] (e.g., kayaking and swimming). These techniques are very useful in analyzing complex movements in greater detail, but such detailed HAR may not be necessary in current healthcare or fitness applications. This is an interesting direction for future research, but it is not in the scope of this paper.

After segmentation, feature extraction is performed, which is aimed to calculate the feature vector of each windowed data. To achieve high accuracy of activity recognition, feature extraction transforms the raw data into numerical features while maintaining the information from the original dataset. Therefore, it produces higher accuracy results than applying ML or DL directly to raw data. Note that this preprocessing stage is commonly found in many artificial intelligence applications to prepare for feature extraction. Following the preprocessing stage, statistical feature extraction is performed before activity recognition starts. Our proposed statistical feature extraction does not require the training phase. It does not require a large volume of data and labeling for training. So, the proposed static feature extraction consumes less computing resources during extraction. Moreover, it can perform the classification of static and dynamic activities within a few seconds. Finally, the proposed statistical feature extraction maintains a higher recognition rate, on average, compared to SVM or DT classifiers.

B. Statistical Feature Extraction

Generally, feature extraction is a common process in HAR, where a feature vector can be extracted at this stage. The extracted features can include information ranging from the time domain to the frequency domain. Based on these features, HAR can get information such as mean, standard deviation, correlation between axes, etc. If feature vectors are available, the proposed HAR method can use the standard deviation values of the x-axis directly. Otherwise, the proposed HAR method has to perform the statistical feature extraction to obtain the x-axis standard deviation values. Using a given feature vector including those values, statistical feature extraction can classify activities as static or dynamic immediately.

Previous work [3] relied on DT classifiers to identify static and dynamic activities, and would then assign a specific classifier (ML or DL) to recognize more specific activities. Another work by Huang et al. [30] utilized an SVM to achieve a similar objective. Both approaches are time-consuming

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

because they involve use of an ML classifier. Instead of using conventional ML classifiers, we use values for standard deviation from the x-axis calculated from an accelerator in the statistical feature extraction stage. The reason we use the standard deviation value from the x-axis is as follows. A closer look into the HAR datasets reveals that inherent statistical information exhibits useful properties for classifying static and dynamic activities. For example, Figs. 2, 3, 4 and 5 show the distributions of standard deviation from the x-axis accelerator data across six activities, based on the commonly used UCI, WISDM, PAMAP, DASDS datasets, respectively. It can be clearly seen that the standard deviation can discriminate the static and dynamic activities in all the HAR datasets that we have examined (UCI, WISDM, PAMAP, and DSADS). This is because static activities do not produce much body movement, so the standard deviation is naturally small; in contrast, the dynamic activities involve huge body movement, so the standard deviation will be larger. For instance, in the UCI dataset in Fig. 2, standing, sitting, and laying (in the red-dotted rectangle) are static activities, which show very different standard deviations compared to dynamic activities including walking, walking downstairs, and walking upstairs (in the blue-dotted rectangle). This implies that by using standard deviation alone, one can classify the activity as static or dynamic on any dataset, which is a more lightweight process compared to using an ML classifier [3], [30]. Hence, we improved the accuracy of the recognition rate in a lightweight manner by just utilizing this observation to extract additional features from the dataset. The distributions along the y-axis and z-axis show similar behavior, but they were omitted from this paper for brevity. This shows that the proposed technique can be used regardless of the wearing position. We denote as $STDEV_X$ the standard deviation of the x-axis value, which can be calculated based on Eq. (1):

$$STDEV_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{1}$$

Algorithm 1 : Statistical Feature Extration

Input : $training\_data$
Output : $training\_data\_for\_static\_activities$, $training\_data\_for\_dynamic\_activities$

Initialize list for $training\_data$, $SigWin$
Sets the size of sampling rate S_r
Calculate the value of $WinLen$ based on sampling rate S_r
Divide $training\_data$ by $subframe$ based on the sampling rate
WHILE (i $< WinLen$){
    Append sliced window $subframe$ to $SigWin$
    Calculate the mean value from the x-axis of each $subframe$
    Calculate $STDEV_x$ from the x-axis of each $subframe$
    IF $STDEV_x$ is $>$ -0.66
        Categorize the activity as dynamic
    ELSE $STDEV_x$ is $\leq$ -0.66
        Categorize the activity as static

The detailed steps of the proposed statistical feature extraction technique are in Algorithm 1. Based on the predefined size of sampling rate S_r, the continuous flow of raw accelerometer data is divided into frames. Each frame is regarded as the same instance of an individual activity. For each frame, HAR can calculate the mean value and the standard deviation. To obtain the standard deviation, we create a new data frame, $SigWin$, then set sampling rate S_r for dividing the training data into smaller windows: the $subframe$. We divide the training data into segments for each $subframe$ based on the sampling rate, then get the total number of subframes, $WinLen$, for all training data. Then, we add each subframe to the new data frame, $SigWin$, based on the total number of subframes, $WinLen$. Finally, we calculate the values for standard deviation and mean from the x-axis of each subframe. If the standard deviation of the x-axis value for the corresponding subframe is greater than -0.66 (obtained from the above explored observation seen in Figs. 2 and 3), the corresponding subframe is classified as a dynamic activity. Otherwise, it is classified as a static activity. Note that we also found the similar results in various open datasets such as UCI-HAR [32], WISDM [33], PAMAP [34], DSADS [35] and UniMiB SHAR [36], wherein the static and dynamic activities can be classified through the proposed technique. However, due to space limitation, we do not report the details in this paper.

After classifying the activity as static or dynamic through the proposed statistical feature extraction technique, further classification is performed with a specific classifier. If statistical feature extraction determines an activity to be static, RF is subsequently used for a more detailed classification of the activity; otherwise, the PQ-CNN is used. The reason we selected RF and the PQ-CNN is that RF shows good performance rather than a DT or another ML classifier for unbalanced datasets as well as balanced datasets. Also, a 1D CNN is commonly used in DL to extract features from time-series sequence data [2], [28].

The proposed HAR method was designed to reduce computations by using a simple RF instead of the computationally heavy CNN for static activities [2], [3], [8], [14], [15], [28], [30], and more importantly, to reduce the storage space required through pruning and quantization with a CNN for dynamic activities. The proposed PQ-CNN achieves a high accuracy rate while reducing computations and storage space. In subsequent discussions, we explain the designs for the RF and PQ-CNN classifiers used in our experiments.

C. Proposed RF Classifier

In the proposed HAR method, RF is used to classify individual static activities. The RF classifier creates multiple decision trees, aggregates the classifications of those trees, and finally classifies them. RF are created to avoid overfitting. To reduce the computational complexity when the user's activity is static, such as laying, sitting, or standing, we use the RF classifier in the proposed HAR method. The highest RF accuracy can be obtained using hyper-parameter tuning that finds an optimal combination of hyper-parameters. The optimal value of a hyper-parameter is described in Section IV.

D. Proposed Pruned and Quantized CNN Model

This section describes the proposed pruned and quantized CNN model for the proposed lightweight HAR method.

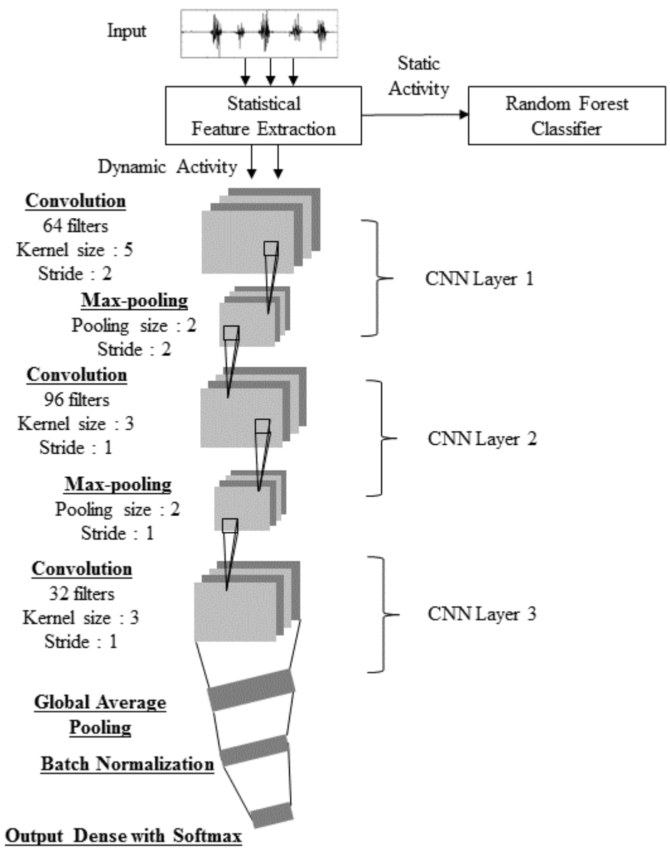

Fig. 6. Frame diagram of the CNN model

Fig. 6 shows the network structure of the CNN: convolution, pooling, batch normalization, and fully-connected layer [2], [28]. Each convolutional and pooling layer are paired, and a batch normalization layer and a fully connected layer follow. This is a commonly used CNN architecture [14], [15], [45] but it can be modified according to the usage of the wearable device. Note that the size of the baseline CNN is still considerably large, which is not suitable for implementation in a wearable device. Quantization for DL can reduce the computation cost of neural network training by replacing low floating-point computations with fast integer computations. Pruning in DL is a method of compression that involves removing weights from a trained model. It can be used for the development a DL model that is smaller and more efficient. To further reduce the network size, we apply pruning and quantization to this baseline CNN. The steps to achieve this PQ-CNN are illustrated in Fig. 7 and further explained in the text. Fig. 7 is the flowchart of the PQ-CNN model. For quantization and pruning for CNN, we use the TensorFlow Model Optimization Toolkit [46]. Data-compression techniques can also be taken into account for additional reduction of the model's size. Since there is often a considerable jump in the compression ratio after a model's weights undergo 16-bit quantization, the pre-trained CNN model is compressed first; then, the process of post-training float16 quantization is performed. We fine-tune the model by applying a pruning API. For pruning, we initially set the model to be 50% sparse, getting increasingly sparser to eventually reach 80%. Finally, we repeat the process of post-training float16 quantization.

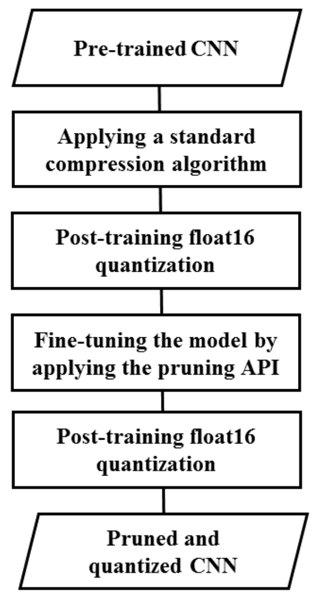

Fig. 7. Flowchart of the PQ-CNN model

IV. Experimental Results and Analysis

This section describes the experimental results and analysis for the proposed lightweight HAR method.

A. Description of the UCI and WISDM Datasets

TABLE II
Distribution of all activities from the UCI and WISDM datasets

Activity	UCI-HAR	WISDM
Sitting	18.7%	1.0%
Laying	19.1%	0%
Standing	17.5%	1.3%
Waking Upstairs	13.4%	10.2%
Waking Downstairs	14.6%	9.7%
Walking	16.7%	40.0%
Jogging	0%	37.7%

TABLE III
Accuracy and processing times for statistical feature extraction with the UCI and WISDM datasets

	UCI	WISDM	
	Accuracy	Accuracy	Processing Time
SVM	99.9%	84.5%	6.6 sec.
DT	99.8%	75.3%	43.5 sec.
Statistical preprocessing	99.7%	99.5%	1.32 sec.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

TABLE IV
OPTIMAL VALUES FOR RF HYPER-PARAMETERS IN THE PROPOSED HAR METHOD

Hyper-parameter	UCI	WISDM
Number of trees	300	200
Max number of levels in each DT	6	6
Min number of data points in a node	20	8
Min number of data points allocated to a leaf node	8	8

TABLE V
AVERAGE ACCURACY COMPARISON OF DIFFERENT CONVENTIONAL ML METHODS WITH THE UCI DATASET

Method	The Proposed HAR Method	
	Static	Dynamic
Logistic Regression	88.27%	94.12%
SVM	87.42%	95.55%
Decision Tree	84.27%	82.30%
Random Forest	**90.80%**	89.75%

TABLE VI
AVERAGE ACCURACY COMPARISON OF DIFFERENT CONVENTIONAL ML METHODS WITH THE WISDM DATASET

Method	The Proposed HAR Method	
	Static	Dynamic
Logistic Regression	87.5%	49.8%
SVM	85.4%	76.3%
Decision Tree	88.9%	61.5%
Random Forest	**92.5%**	74.2%

We chose the UCI [32] and WISDM [33] datasets, which are widely used public datasets for training and performance evaluation of HAR methods. The reason for choosing these datasets is because they are open to public access, and they are also widely used in many previous publications in the field of HAR [9], [14], [15], [28], [30]. The UCI dataset was collected from 30 subjects performing six different activities: laying, sitting, standing, walking, walking upstairs, and walking downstairs. Each person carried a smartphone on the waist and performed all six activities. The sensor signals were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 second and 50% overlap. The experiments were recorded on video to manually label the data later. The obtained dataset was randomly partitioned into two parts: 70% for training data, and 30% for test data. The WISDM dataset consists of 1,098,207 recordings from the tri-axial accelerometers of 36 users performing six activities: jogging, sitting, standing, walking, walking upstairs, and walking downstairs. The data was collected using an accelerometer sensor embedded in a smart phone with a sampling rate of 20 Hz. The WISDM dataset includes six columns for user ID, activity label, time stamp, and accelerometer readings in the x, y, and z directions. As shown in Table II, we can see that the UCI dataset is balanced for all activities. The WISDM dataset is imbalanced, because there is a lot of walking and jogging activity, but far less sitting and standing. All the experiments were conducted on a system equipped with an Intel Core i7-6700 CPU and 8GB RAM.

TABLE VII
AVERAGE ACCURACY FROM THE PROPOSED HAR METHOD FOR DYNAMIC ACTIVITIES WITH THE UCI DATASET

The Proposed DL Method	Accuracy	Model Size
Baseline CNN	96.32%	148.96 KB
Baseline CNN (TFlite)	96.32%	74.90 KB
Pruned CNN	95.89%	52.26 KB
Quantized and pruned CNN	**95.89%**	**30.76 KB**

TABLE VIII
AVERAGE ACCURACY FROM THE PROPOSED HAR METHOD FOR DYNAMIC ACTIVITIES WITH THE WISDM DATASET

The Proposed DL Method	Accuracy	Model Size
Baseline CNN	96.83%	227.99 KB
Baseline CNN (TFlite)	96.83%	113.82 KB
Pruned CNN	95.10%	76.45 KB
Quantized and pruned CNN	**95.10%**	**45.57 KB**

B. Performance Evaluation for Statistical Feature Extraction

In order to minimize computational complexity, the proposed HAR method classifies static and dynamic activities by using the proposed statistical feature extraction. Table III shows the accuracy and processing times for statistical feature extraction by DT and SVM classifiers with the UCI and WISDM datasets. The recognition rate of the proposed statistical feature extraction technique when classifying dynamic or static activities is similar to those from DT and SVM classifiers with the UCI dataset. However, outstanding performance was obtained with the WISDM dataset, which was 15% and 24.2% more accurate than the SVM and DT classifiers, respectively. Moreover, statistical feature extraction only requires a very small amount of computation, compared to the SVM and DT classifiers. With the WISDM dataset, total latency in statistical feature extraction was only 1.325 seconds when classifying 1,098,204 activities, which is 5 times and 32.95 times faster than the SVM and DT classifiers, respectively. Therefore, the preprocessing time to classify static and dynamic activities was negligible, compared to the DT or SVM classifiers.

Since the UCI dataset provides an additional 561 feature vectors (including standard deviation), the statistical feature extraction method does not need to calculate the standard deviation from the x-axis. Using the given feature vectors, therefore, the proposed statistical feature extraction can classify activities as static or dynamic immediately.

C. Performance Evaluation for RF Classification

Hyper-parameter optimization is important because it makes a significant impact on the learning algorithm's performance. The performance from ML and DL algorithms depends to a

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

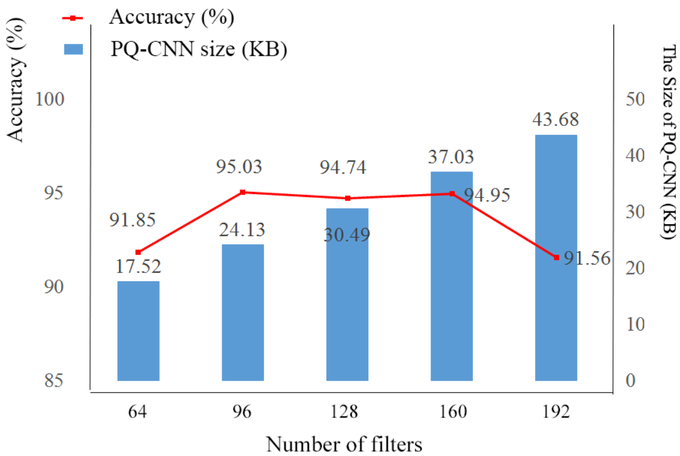

Fig. 8. Performance of the PQ-CNN model from increasing the number of filters at the convolutional layer with the UCI dataset

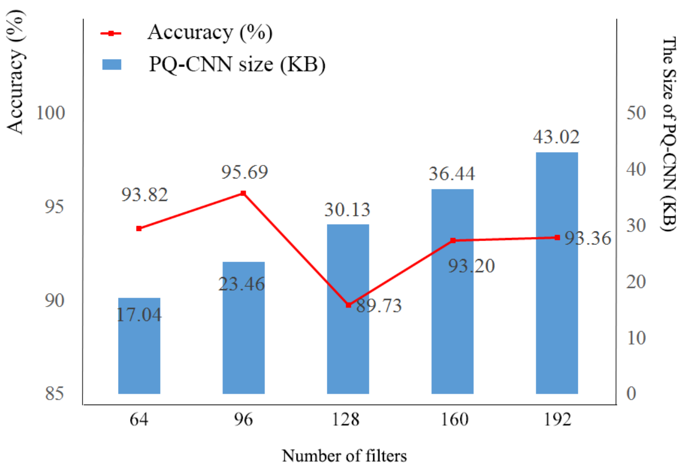

Fig. 9. Performance of the PQ-CNN model with from increasing the number of filters at the convolutional layer with the WISDM dataset

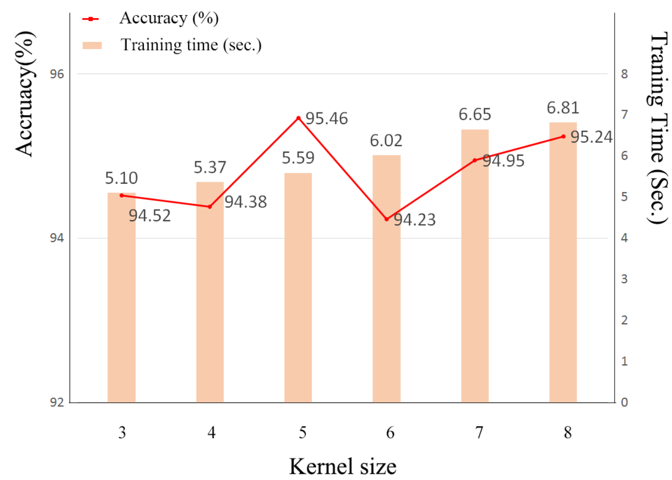

Fig. 10. Performance of the PQ-CNN model from increasing the kernel size at the convolutional layer when classifying the UCI dataset

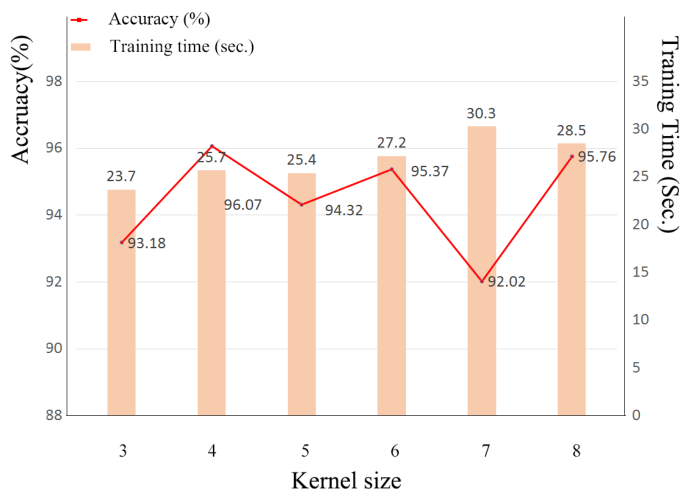

Fig. 11. Performance of the PQ-CNN model from increasing the kernel size at the convolutional layer when classifying the WISDM dataset

large extent on the value of the hyper-parameter. We performed five-fold cross-validation on all datasets. For cross-validation, we split the UCI and WISDM datasets into training and validation sets repeatedly. Based on this way, we selected our optimal hyperparameters value. Table IV shows the optimal values of hyper-parameters for RF in the proposed HAR method when using the UCI and WISDM datasets. We can use the GridSearchCV function provided by the Scikit-Learn wrapper in the Keras API to automate the tuning of hyperparameters of the RF classifier in any HAR datasets. [46]. Based on the values of the parameters, we determined the accuracy of the proposed HAR method using RF. From the above optimal parameter values, accuracy rates of 90.80% and 92.5% for classification of static activities using the RF classifier were achieved with the UCI and WISDM datasets, respectively. In practical applications, the hyperparameters are usually stored on a non-volatile memory (e.g., NAND flash) and loaded to the RAM when the HAR takes place. During the production time, we can store multiple sets of hyperparameters that are optimal to different datasets on the non-volatile memory, so that the users can choose them based on different use cases. On top of that, we can also allow the users to perform firmware updates from time to time, so that they can obtain new sets of RF hyperparameters that are optimal for new datasets.

Tables V and VI show the overall performance from RF in the proposed HAR method with the UCI and WISDM datasets, respectively. In Table V, the RF accuracy for static activities with the UCI dataset was 90.8%, denoted in bold typeface, outperforming all the other approaches. In Table VI, the RF accuracy for static activities was 92.5%, also denoted in bold, attaining the best HAR rate. Hence, it was confirmed that the RF classifier performed well when recognizing human static activity with UCI as well as WISDM datasets. RF showed stable and good performance with the unbalanced dataset as well as the balanced dataset.

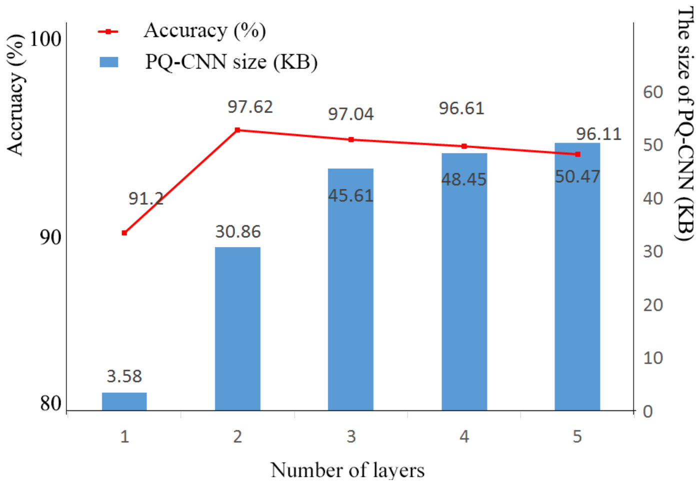

Fig. 12. Performance of the PQ-CNN model from increasing the number of layers while classifying the UCI dataset

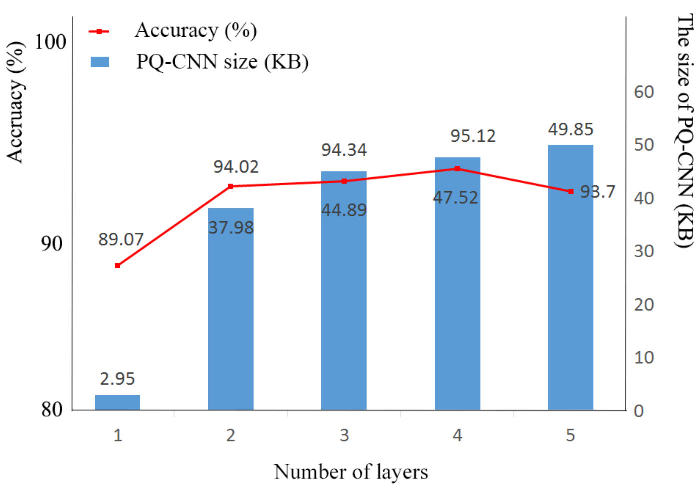

Fig. 13. Performance of the PQ-CNN model from increasing the number of layers while classifying the WISDM dataset

D. *Performance Evaluation of the PQ-CNN*

This section describes the performance analysis of the PQ-CNN in terms of model size and accuracy in HAR. Then, we present the impact of hyper-parameters on the PQ-CNN. From these experiment results, we observed that loss of the PQ-CNN model's accuracy was not significant (less than 1%), but the size of the CNN model dropped by about five times. Based on the result of the impact of hyper-parameters on the PQ-CNN, we can select the optimal parameters for the model.

Tables VII and VIII show the results of the quantized and pruned CNN with the original CNN model on the UCI and WISDM datasets, respectively. For the UCI dataset, as shown in Table VII, the loss of recognition accuracy after pruning and quantization was reduced by 0.43%, which is not significant. However, the size of the PQ-CNN model was reduced by 4.84 times. Similarly, for the WISDM dataset (Table VIII), the proposed PQ-CNN model showed similar accuracy, compared to the baseline CNN model, with only a

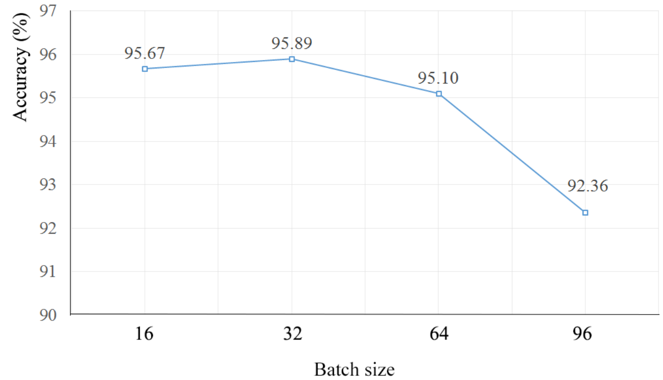

Fig. 14. Performance of the PQ-CNN model based on batch size with the UCI dataset

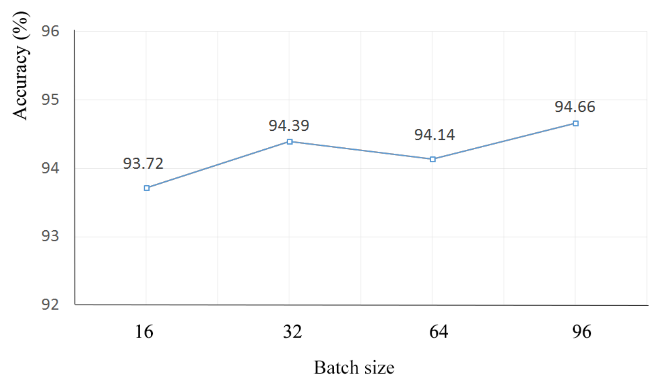

Fig. 15. Performance of the PQ-CNN model based on batch size with the WISDM dataset

1.73% loss of accuracy, but the size was reduced by 5 times. From these results, we can see that the proposed PQ-CNN can greatly reduce the size of the CNN model by nearly five times while maintaining accuracy, compared to the baseline CNN model. Below, we present the effect of hyper-parameters on the PQ-CNN, such as the effect from the number of filters, from the filter size, from the number of layers, and from the batch size. Based on the analysis of the PQ-CNN, optimal parameters for the model were selected.

1) Effect of the number of filters: CNNs have several different filters/kernels that can convolve the given training data spatially to detect features. As the number of filters in the convolutional layer of the CNN increases, the depth of the feature space increases. So, a large number of filters in a CNN helps to learn more levels of global abstract structures. Figs. 8 and 9 show the accuracy and size of the PQ-CNN model based on changing the number of filters at the second convolutional layer. The left axis denotes the accuracy while the right axis represents the size of the PQ-CNN model concerning the increased number of filters in the PQ-CNN model, respectively. We cannot say the accuracy of the model increases with an increased number of filters. The best accuracy was obtained from 96 filters in convolutional layers when classifying the UCI and WISDM datasets

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3266506

TABLE IX
CLASSIFICATION CONFUSION MATRIX FOR THE UCI DATASET

		Predicted Class						Recall (%)
		Walking	Walking Upstairs	Walking Downstairs	Sitting	Standing	Laying	
Actual Class	Walking	**480**	1	8	0	0	0	98.16
	Walking Upstairs	0	**460**	10	0	0	0	97.87
	Walking Downstairs	2	10	**420**	0	1	0	97.00
	Sitting	0	0	0	**415**	62	0	87.00
	Standing	0	1	0	80	**466**	0	85.19
	Laying	0	0	0	0	0	**531**	100.00
Precision (%)		99.59	97.46	95.89	83.84	88.09	100.00	**94.06**

TABLE X
CLASSIFICATION CONFUSION MATRIX FOR THE WISDM DATASET

		Predicted Class						Recall (%)
		Walking Downstairs	Jogging	Walking Upstairs	Walking	Sitting	Standing	
Actual Class	Walking Downstairs	**595**	4	5	6	0	0	97.54
	Jogging	6	**1970**	18	31	1	1	97.19
	Walking Upstairs	44	1	**571**	6	0	0	91.80
	Walking	24	31	10	**2396**	0	1	97.36
	Sitting	0	1	0	1	**411**	39	91.13
	Standing	1	0	0	0	22	**350**	93.83
Precision (%)		88.81	98.16	94.54	98.20	94.70	89.51	**96.14**

2) Effect of the kernel size: Figs. 10 and 11 show the effect of the kernel size at the second convolutional layer in the PQ-CNN model. The left axis denotes the accuracy while the right axis represents the training time of the PQ-CNN model concerning the increased number of filters in the PQ-CNN model, respectively. When increasing the kernel size, training time for the PQ-CNN increased. We got the best accuracy with 95.46% when the kernel size was 5 while classifying the UCI dataset. But we got better accuracy with 96.07% when the kernel size was 4 while classifying the WISDM dataset. If the kernel size for the PQ-CNN model increases, more features can be extracted. Therefore, the accuracy of the PQ-CNN model can be increased. But we can see that as the training time for the PQ-CNN model increases, so does overfitting.

3) Effect of the number of layers: Figs. 12 and 13 show the effect from increasing the number of convolutional and pooling layers in the PQ-CNN model. The left axis denotes the accuracy while the right axis represents the size of the PQ-CNN model concerning the increased number of filters in the PQ-CNN model, respectively. With an increasing number of convolutional and pooling layers, the size of the model increases. We got the best accuracy with 97.62% when the number of convolutional and pooling layers in the PQ-CNN model was 2 while classifying the UCI dataset. But the best accuracy was only 95.12% when the number of convolutional and pooling layers in the model was 4 while classifying the WISDM dataset. If the number of convolutional and pooling layers in the PQ-CNN model increases, more features can be extracted. Therefore, the accuracy of the PQ-CNN model can be increased, but overfitting cannot be avoided.

4) Effect of the batch size: The mini-batch methodology is a variation of the gradient descent algorithm. It splits the training data into small batches, which are then used for training. Figs. 14 and 15 present classification accuracy with respect to batch sizes while classifying the UCI and WISDM datasets, respectively. We got the highest accuracy with 95.89% when the batch size was 32 with the UCI dataset. The highest accuracy was 94.66% when the batch size was 96 with the WISDM dataset.

E. Overall Performance for the Proposed HAR

Tables IX and X show the confusion matrixes from the proposed lightweight HAR method when classifying the UCI and WISDM datasets, respectively. An F1 score for the model can be calculated using Eqs. (2)-(5).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

(TP : True Positive, TN : True Negative, FP : False Positive, FN : False Negative)

From Tables IX and X, we got F1 scores of 0.9417 and 0.9438 with the UCI and WISDM datasets, respectively. We observe that the accuracies for sitting and standing classifications were lower than the other activity classes, as shown in Table XI, which is a result found in previous work [40], [47]. Also, we can see that the accuracies for standing and walking downstairs classifications were lower than the other activity classes, as shown in Table X. However, this is a result different from previous work owing to the properties of an unbalanced dataset.

Tables XI and XII show overall performance from the proposed HAR method compared with several previously

TABLE XI
AVERAGE ACCURACY OF THE DIFFERENT ML & DL METHODS WITH THE UCI DATASET

Method	F1 score	Model	Sensor	Size of Model (KB)
Logistic Regression [47]	0.9370	-	**A+G**	
SVM [47]	0.9430	-	**A+G**	
Decision Tree [47]	0.8500	-	**A+G**	
Random Forest [47]	0.9210	-	**A+G**	
Teng *et al.* [14]	0.9697	CNN	**A+G**	1433.6
Tang *et al.* [15]	0.9627	CNN	**A+G**	1331.2
Gao *et al.* [9]	0.9721	CNN	**A+G**	1843.2
Xia *et al.* [40]	0.9578	LSTM + CNN	**A**	204.8
The Proposed 2-level HAR	**0.9427**	RF + CNN	**A**	764.0
The Proposed 2-level HAR	**0.9417**	RF + PQ-CNN	**A**	645.8

A : Accelerometer, **G** : Gyroscope

published results that used ML and DL techniques with the UCI and WISDM datasets, respectively. To evaluate the impact of pruning and quantization on the classification accuracy in the proposed lightweight HAR method, the combination of the RF and the baseline CNN (RF+CNN) and the combination of RF and PQ-CNN (RF+PQ-CNN) were compared in UCI-HAR and WISDM datasets, respectively. Similar to [9], [14], [15], [40], adding more layers or filters to the DL model helps extract more features. Moreover, the combination of several different predictions from different models can result in better predictions than from any single model [40]. But the superior accuracy of DL models comes at the cost of high computational complexity. The processing time of a DL classifier increases as the number of convolutional layers in the DL classifier increases. Using a larger DL model and more sensors increases the computational complexity and memory size significantly. Therefore, we can see again that there is a trade-off between complexity and accuracy in the DL model. We can see that the proposed lightweight HAR method is slightly less accurate, compared to [9], [14], [15], [40], but the model size is much smaller. The study in [40] showed a smaller model size compared to our proposed HAR method, but we note that LSTM is computationally heavy compared to the combination of RF and PQ-CNN classifiers. In the proposed lightweight HAR method, the size of the RF model was 615 KB and 55 KB for the UCI and WISDM datasets, respectively, whereas the size of the PQ-CNN was 30.76 KB and 45.57 KB for the UCI and WISDM datasets, respectively. The size of the RF model is larger than the PQ-CNN because the number of trees in RF depends on the number of rows in the dataset. But RF is significantly lightweight, compared to a CNN in terms of computational complexity [49].

F. Discussion

The proposed lightweight HAR is more advantageous compared to the existing works [9], [14], [15], [40]. STM32 microcontroller [5] is one of the representative hardware platforms to implement wearable devices with only 1MB SRAM. As shown in Table XI and XII, the existing HAR methods based on CNN require a memory size of over 1.3 MB, which makes them hard to implement in wearable devices. However, the proposed lightweight HAR method only consumes 645.8 KB (UCI) and 100.6 KB (WISDM), so it can be directly implemented on an STM32 microcontroller [5]. Given that wearable devices are worn for long period to perform continuous monitoring, we think that it is more important to develop a HAR method with moderately high accuracy and lightweight computation, rather than high accuracy with heavy computation. This is because heavy computation can drain up the battery quickly, thus defeating long-period monitoring using in wearable devices. The proposed RF and PQ-CNN are computationally lightweight, so they can be executed on this microcontroller with only 80 MHz clocks. Moreover, the proposed lightweight HAR only utilizes data from one accelerometer, which further reduces the cost and energy consumption of the wearable device. Existing DL framework (e.g., TensorFlow Lite) [50] also provides the facility to convert the DL model into microcontroller-friendly C codes for practical deployment. In contrast, the HAR method based on DL [9], [14], [15] requires a huge memory size, which is not supported by the common microcontroller used to build wearable devices. On the other hand, the solution based on LSTM + CNN [40] has a small size model, but the computation is heavier compared to RF + PQ-CNN used in our HAR method. Hence, the proposed HAR method is more wearable-ready than compares to existing works [9], [14], [15], [40].

Using the frequently activated lightweight ML and single accelerometer, the proposed hybrid HAR method is simple and practical. Moreover, the proposed HAR method achieves the equivalent level of accuracy even after pruning and quantization, and performs well regardless of dataset, balanced or unbalanced. The proposed compact can continuously recognize and track human activity over a long period of time, therefore it will be useful in scenarios such as when designing tiny wearable rescue working under extreme constraints of battery usage, even in temperatures of minus 20 degrees [51]. It can also be utilized for e-skin sensor for patient fall prevention when continuous monitoring is more important than accurate measurement [52].

V. CONCLUSION

In this paper, we proposed an optimal HAR method to wearable devices based on statistical feature extraction combined with RF and PQ-CNN classifiers using data obtained from a single accelerometer. Statistical feature extraction drastically reduces computational complexity in the first level of the proposed two-level HAR method. Combing RF and PQ-CNN classifiers achieves both higher accuracy and lower complexity

TABLE XII
AVERAGE ACCURACY OF THE DIFFERENT ML & DL METHODS WITH THE WISDM DATASET

Method	F1 score	Model	Sensor	Size of Model (KB)
Logistic Regression [48]	0.8110	-	**A**	
SVM [48]	0.8620	-	**A**	
Decision Tree [48]	0.8745	-	**A**	
Random Forest [48]	0.8210	-	**A**	
Teng *et al.* [14]	0.9881	CNN	**A**	10649.6
Tang *et al.* [15]	0.9751	CNN	**A**	1679.3
Gao *et al.* [9]	0.9813	CNN	**A**	1474.6
Xia *et al.* [40]	0.9585	LSTM + CNN	**A**	204.8
The Proposed 2-level HAR	**0.9454**	RF + CNN	**A**	283.0
The Proposed 2-level HAR	**0.9438**	RF + PQ-CNN	**A**	100.6

A : Accelerometer

(computation, memory footprint) in the second level compared to state-of-the-art approaches. A single sensor makes the proposed method more practical when deployed in small size, resource-constraint, low-cost wearable devices to perform continuous tasks.

In future, we plan to extend our work on recognize high-level activities, which are much more complex, but semantically consistent with a human real life. Most of the successful existing HAR approaches have relied on the DL model trained on specific persons or specific positions. However, these ML or DL models do not perform well in different persons and positions. Therefore, we plan to investigate a new DL model to overcome this challenge of the cross-domain HAR problem. On the other hand, we also plan to explore the detailed segmentation of specific activities (e.g., swimming), which can provide more insights into HAR research. We will continue to explore other feature extraction techniques that can be useful for wearable devices.

REFERENCES

[1] Binh Nguyena, Yves Coelho, Teodiano Bastos, and SridharKrishnana, "Trends in human activity recognition with focus on machine learning and power requirements," *Machine Learning with Applications*, Vol 5(15) September 2021.

[2] T. Zebin, P. Scully, N. Peek, A. Casson, and K. Ozanyan, "Design and implementation of a convolutional neural network on an edge computing smartphone for human activity recognition," *IEEE Access*, vol. 7, pp. 133509-133520, 2019.

[3] Yves Luduvico Coelho , Francisco de Assis Souza dos Santos, Anselmo Frizera-Neto , and Teodiano Freire Bastos-Filho, "A Lightweight Framework for Human Activity Recognition on Wearable Devices," *IEEE Sensors Journal*, Vol. 21(21), 2021.

[4] "Galaxy S22," [Online] Available: https://www.samsung.com/levant/smartphones/galaxy-s22/

[5] "STM32 Nucleo-144 development board with STM32H755ZI MCU," [Online] Available: https://www.st.com/en/evaluation-tools/nucleo-h755zi-q.html

[6] Vijay Bhaskar Semwal, Praveen Lalwani, Manas Kumar Mishra, Vishwanath Bijalwan, and Jasroop Singh Chadha, "An optimized feature selection using bio-geography optimization technique for human walking activities recognition", *Computing*, pp. 2893–2914, 2021.

[7] Anjali Gupta and Vijay Bhaskar Semwal,"Multiple task human gait analysis and identification: ensemble learning approach", Emotion and information processing. Springer, pp. 185-197, 2020.

[8] Wenbo Huang, Lei Zhang, Qi Teng, Chaoda Song, and Jun He, "The Convolutional Neural Networks Training With Channel-Selectivity for Human Activity Recognition Based on Sensors", *IEEE Journal of Biomedical and Health Informatics*, Vol.25(10), pp. 3834 - 3843, Oct. 2021

[9] Wenbin Gao, Lei Zhang, Wenbo Huang, Fuhong Min, Jun He , and Aiguo Song., "Deep Neural Networks for Sensor-Based Human Activity Recognition Using Selective Kernel Convolution," *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*, VOL. 70, 2021.

[10] G. Ascioglu and Y. Senol, "Design of a Wearable Wireless Multi-Sensor Monitoring System and Application for Activity Recognition Using Deep Learning",*IEEE Access*, vol. 8, pp.169183-169195, 2020.

[11] Yonatan Vaizman, Katherine Ellis, and Gert Lanckriet, "Recognizing Detailed Human Context in the Wild from Smartphones and Smartwatches," *IEEE Pervasive Computing*, Vol.16(4), pp. 62 - 74, October 2017.

[12] B. Zhou, H. Wang, F. Hu, N. Feng and H. Tang, "Accurate recognition of lower limb ambulation mode based on surface electromyography and motion data using machine learning", Computer Methods and Programs in Biomedicine, vol. 193, pp. 105486(1-30), 2020.

[13] Meng-Chieh Yu at al, "Big Data Small Footprint : The Design of A Low-Power Classifier for Detecting Transportation Modes," Proceedings of the VLDB, Vol. 7, No. 13, pp. 1429-1440, 2014

[14] Qi Teng, Kun Wang, Lei Zhang, and Jun He, "The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition,"*IEEE Sensors J.*, vol. 20, no. 13, pp. 7265–7274, Jul. 2020.

[15] Yin Tang, Qi Teng, Lei Zhang, Fuhong Min, and Jun He, "Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors," *IEEE Sensors J.*, vol. 21(1), pp. 581–592, Jan. 2021.

[16] Rahul Jain, Vijay Bhaskar Semwal, and Praveen Kaushik, "Stride segmentation of inertial sensor data using statistical methods for different walking activities", *Robotica*, Vol 10(8), pp. 1-14, 2021

[17] Anjali Gupta and Vijay Bhaskar Semwal, "Occluded Gait reconstruction in multi person Gait environment using different numerical methods. *Multimedia Tools and Applications*, Vol. 81, pp. 23421–23448, 2022.

[18] Vijay Bhaskar Semwal, Arghya Mazumdar, Ashish Jha, Neha Gaud, and Vishwanath Bijalwan, "Speed, cloth and pose invariant gait recognition-based person identification", Machine learning: theoretical foundations and practical applications, Springer, 39-56, 2021.

[19] Prithvi Patil, K.Shusheel Kumar, Neha Gaud, and Vijay Bhaskar Semwal, "Clinical Human Gait Classification: Extreme Learning Machine Approach", 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019.

[20] Vijay Bhaskar Semwal, Neha Gaud, Praveen Lalwani, Vishwanath Bijalwan, and Abhay Kumar Alok, "Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor", *Artificial Intelligence Review*, Vol. 55, pp.1149–1169, 2022

[21] Semwal, Vijay Bhaskar, Anjali Gupta, and Praveen Lalwani, "An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition", *The Journal of Supercomputing*, pp 12256–12279, 2021

[22] Rahul Jain,Vijay Bhaskar Semwal, and Praveen Kaushik, "Deep ensemble learning approach for lower extremity activities recognition using wearable sensors", *Expert Systems*, Vol 39(6), 2021.

[23] Challa, Sravan Kumar, Akhilesh Kumar, and Vijay Bhaskar Semwal, "A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data",*The Visual Computer*, 1-15, 2021.

[24] Nidhi Dua, Shiva Nand Singh, Vijay Bhaskar Semwal, and Sravan

Kumar Challa, "Inception inspired CNN-GRU hybrid network for human activity recognition", *Multimedia Tools and Applications*, 1-35, 2022.

[25] Nidhi Dua, Shiva Nand Singh, and Vijay Bhaskar Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors." *Computing*, Vol. 103, pp. 1461–1478, 2021.

[26] Vishwanath Bijalwan, Vijay Bhaskar Semwal, Ghanapriya Singh, and Tapan Kumar Mandal, "HDL-PSR: Modelling Spatio-Temporal Features using Hybrid Deep Learning Approach for Post-Stroke Rehabilitation", *Neural Processing Letters*, 2022.

[27] Vishwanath Bijalwan, Vijay Bhaskar Semwal, and Vishal Gupta, "Wearable sensor-based pattern mining for human activity recognition: deep learning approach", *Industrial Robot*, Vol 49(2), pp.21-33, 2021.

[28] H. Cho and S. Min Yoon, "Divide and Conquer-Based 1D CNN Human Activity Recognition Using Test Data Sharpening," *Sensors*, Vol. 18(4), April 2018.

[29] Ganapati Bhat, Yigit Tuncel, Sizhe An, Hyung Gyu Lee, and Umit Y. Ogras, "An Ultra-Low Energy Human Activity Recognition Accelerator for Wearable Health Applications," *ACM Transactions on Embedded Computing Systems*, Vol. 18(5), pp 1–22, 2019.

[30] Jiahui Huang, Shuisheng Lin, Ning Wang, Guanghai Dai, Yuxiang Xie, and Jun Zhou, "TSE-CNN: A Two-Stage End-to-End CNN for Human Activity Recognition," *IEEE Journal of Biomedical and Health Informatics*, Vol. 24(1). 2019.

[31] US Department of Labor. 2017. American Time Use Survey. [Online] https://www.bls.gov/tus/, accessed 13 July 2019.

[32] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz, " A Public Domain Dataset for Human Activity Recognition Using Smartphones", 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

[33] Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore, "Activity Recognition using Cell Phone Accelerometers", ACM SigKDD Explorations Newsletter, Vol. 12, No. 2, pp. 74–82, 2011.

[34] A. Reiss and D. Stricker, "Introducing a New Benchmarked Dataset for Activity Monitoring", The 16th IEEE International Symposium on Wearable Computers(ISWC), 2012.

[35] B. Barshan and M. C. Yüksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units", *The Computer Journal*, Voo. 57(11), pp.1649–1667, November 2014.

[36] Micucci, D., Mobilio, M., and Napoletano, P., "UniMiB SHAR: A new dataset for human activity recognition using acceleration data from smartphones", *IEEE Sens. Lett.*, Vol. 2, pp. 15-18, 2016.

[37] J. Cervantes at al., "SVM Classification for Large Data Sets by Considering Models of Classes Distribution", Sixth Mexican International Conference on Artificial Intelligence, Special Session (MICAI), 2007.

[38] Hasmat Malik, Nuzhat Fatema, and Atif Iqbal, "Intelligent Data-Analytics for Condition Monitoring : Smart Grid Applications", Elsevier, pp.215-241, 2021.

[39] Omer Sagi,Lior Rokach, "Ensemble learning: A survey", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, pp. 1-18, 2018

[40] Kun Xia, Jianguang Huang, and Hanyu Wang, "LSTM-CNN Architecture for Human Activity Recognition,*IEEE Access*, Vol. 8, pp: 56855 - 56866, March 2020.

[41] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac, and Parisa Rashidi, "Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey,*IEEE Access*, Vol. 8, pp: 210816 - 210836, November 2020.

[42] Jun-Huai Li, Ling Tian, Huaijun Wang, Yang An, and Kan Wang,Lei Yu, "Segmentation and Recognition of Basic and Transitional Activities for Continuous Physical Human Activity," *IEEE Access*, Vol. 7, pp. 42565 - 42576, March 2019.

[43] Sen Qiu, Zhengdong Hao, Zhelong Wang, Long Liu, Jiayi Liu, Hongyu Zhao, Giancarlo Fortino, "Sensor Combination Selection Strategy for Kayak Cycle Phase Segmentation Based on Body Sensor Networks", *IEEE Internet of Things Journal*, Vol 9(6), 2022

[44] Jiaxin Wang, Zhelong Wang, Fengshan Gao, Hongyu Zhao, Sen Qiu, and Jie Li, "Swimming Stroke Phase Segmentation Based on Wearable Motion Capture Technique", *IEEE Transactions on Instrumentation and Measurement*, Vol. 69(10), 2020

[45] Z. Yang, O. I. Raymond, C. Zhang, Y. Wan, and J. Long, "DFTerNet: Towards 2-bit dynamic fusion networks for accurate human activity recognition," *IEEE Access*, vol. 6, pp. 56750–56764, 2018.

[46] F. Chollet. (2013) Keras: The python deep learning library. [Online]. Available: https://keras.io/

[47] Weiheng Kong, Lili He, and Hailong Wang, "Exploratory Data Analysis of Human Activity Recognition Based on Smart Phone", *IEEE Access*, Vol.9, pp.73355-73364, 2021.

[48] Shivanjali Khare, Sayani Sarkar, and Michael Totaro, "Comparison of Sensor-Based Datasets for Human Activity Recognition in Wearable IoT", 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), October 2020

[49] EvaGarcía-Martína, Crefeda FaviolaRodrigues, GrahamRiley, and HåkanGrahna, "Estimation of energy consumption in machine learning", *Journal of Parallel and Distributed Computing*, Vol.134, pp. 75-88, December 2019.

[50] "TensorFlow Lite for Microcontrollers," [Online] Available: https://www.tensorflow.org/lite/microcontrollers

[51] Cristian Culman, Samaneh Aminikhanghahi, and Diane J. Cook, "Easing Power Consumption of Wearable Activity Monitoring with Change Point Detection," *Sensors (Basel)*. 20(1), 310, 2020.

[52] A.-C. Bunea, V. Dediu, E.A Laszlo, F. Pistrițu, M. Carp, , F.S. Iliescu, O. N. Ionescu, C. Iliescu, "E-Skin: The Dawn of a New Era of On-Body Monitoring Systems," *Micromachines*, Vo. 12(9), 1091, 2021

MYUNG-KYU YI (Member, IEEE) received a Ph.D. in Computer Science and Engineering from Korea University in 2005. He is currently a Research Professor with Gachon University, and is a member of the personal health information standardization task force of the TTA U-Health project group. His research interests include healthcare, security, machine learning, deep learning, and human activity recognition.

WAI-KONG LEE (Member, IEEE) received a B.Eng. in electronics and an M.Eng.Sc. from Multimedia University in 2006 and 2009, respectively, with a Ph.D. in engineering conferred by University Tunku Abdul Rahman (UTAR), Malaysia, in 2018. From 2009 to 2012, he served as a Research and Development Engineer for several multinational companies, including Agilent Technologies (now known as Keysight) Malaysia. He served as an Assistant Professor and as Deputy Dean (research and development) for the Faculty of Information and Communication Technology at UTAR. He was a Visiting Scholar at Carleton University, Canada, in 2017, at Feng Chia University, Taiwan, in 2016 and 2018, and at OTH Regensburg, Germany, in 2015, 2018, and 2019. He is currently a Postdoctoral Researcher at Gachon University, South Korea. His research interests include cryptography, GPU computing, numerical algorithms, the Internet of Things (IoT), and energy harvesting.

SEONG OUN HWANG (Senior Member, IEEE) received the B.S. degree in mathematics from Seoul National University, in 1993, the M.S. degree in information and communications engi- neering from the Pohang University of Science and Technology, in 1998, and the Ph.D. degree in computer science from the Korea Advanced Institute of Science and Technology, South Korea. He worked as a Software Engineer with LG- CNS Systems, Inc., from 1994 to 1996. He also worked as a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), from 1998 to 2007. He worked as a Professor with the Department of Software and Communications Engineering, Hongik University, from 2008 to 2019. He is currently a Professor with the Department of Computer Engineering, Gachon University. His research interests include cryptography, cybersecurity, and artificial intelligence. He is an Editor of ETRI Journal.