IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 25, 2022, accepted May 4, 2022, date of publication May 10, 2022, date of current version May 17, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174124

ResNet-SE: Channel Attention-Based Deep Residual Network for Complex Activity Recognition Using Wrist-Worn Wearable Sensors

SAKORN MEKRUKSAVANICH[1], (Member, IEEE), ANUCHIT JITPATTANAKUL[2], KANOKWAN SITTHITHAKERNGKIET[2], PHICHAI YOUPLAO[3], AND PREECHA YUPAPIN[4]

[1]Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand
[2]Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
[3]Department of Electrical Engineering, Faculty of Industry and Technology, Rajamangala University of Technology Isan Sakon Nakhon Campus, Sakon Nakhon 47160, Thailand
[4]Department of Electrical Technology, Faculty of Industrial Technology, Institute of Vocational Education Northeastern 2, Sakon Nakhon 47000, Thailand

Corresponding author: Phichai Youplao (phichai.yo@rmuti.ac.th)

This work was supported in part by the Thailand Science Research and Innovation Fund; in part by the University of Phayao, under Grant FF65-RIM041; in part by the National Science, Research and Innovation Fund (NSRF); and in part by the King Mongkut's University of Technology North Bangkok under Contract KMUTNB-FF-65-27.

ABSTRACT Smart mobile devices are being widely used to identify and track human behaviors in simple and complex daily activities. The evolution of wearable sensing technologies pertaining to wellness, living surveillance, and fitness tracking is based on the accurate analysis of people's behavior from the data acquired through different sensors embedded in smart devices, especially wrist-worn wearable technologies such as smartwatches. Many deep learning techniques have been developed to realize human activity recognition (HAR), with simple daily activities being focused on. However, several challenges remain to be addressed in complex HAR research involving specific human behaviors in different contexts. To address the problems pertaining to complex HAR, a deep neural network composed of convolutional layers and residual networks was developed in this work. Additional attention was incorporated in the system by using a squeeze-and-excite mechanism. The model effectiveness was investigated considering three publicly available datasets, (WISDM-HARB, UT-Smoke, and UT-Complex). The proposed network achieved overall accuracies of 94.91%, 98.75%, and 97.73% over WISDM-HARB, UT-Smoke, and UT-Complex, respectively. The results showed that deep residual networks are more durable and superior at activity recognition than the existing models.

INDEX TERMS Wrist-worn wearable sensor, deep learning, deep residual network, attention mechanism, complex activity recognition.

I. INTRODUCTION

Wearable technologies refer to a general-purpose computing frameworks with multiple sensors that enable real-time monitoring of human activities in various domains, such as healthcare, sport and exercise monitoring, and inappropriate behavior prevention [1], [2]. For instance, in healthcare systems, identification of physical activity based on wrist-worn wearable sensor information can help avoid adverse consequences associated with poor lifestyle choices. For example, monitoring the amount of time a person spends in eating-related behaviors can facilitate the treatment of conditions such as obesity, diabetes, cancer, and cardiovascular diseases [3]. Another example is smoking detection, through which an individual or a healthcare professional can help people limit their smoking by gaining a better understanding of their everyday smoking habits [4].

Smartwatches, which are widely available and affordable in the present markets, and being increasingly used by people in their everyday lives. Most smartwatches have inertial

The associate editor coordinating the review of this manuscript and approving it for publication was Razi Iqbal.

measurement unit (IMU) sensors, such as accelerometers, gyroscopes, and magnetometers [5]. Consequently, smartwatches are highly personalized because they can be worn or carried at all times and used for various reasons, such as for tracking eating or typing habits [6]. The computational capacity and versatility of these devices is increasing, and the cost, scale, and energy consumption is decreasing. Wearing a smartwatch is a safe approach for movement detection since it eliminates the need to wear several sensors in various locations, which may be inconvenient for the elderly and patients [7], [8]. Additionally, a smartwatch is a superior alternative than a smartphone because people wear smartwatches in the daytime and smartphones are often not as easy to reach as smartwatches, particularly when sleeping or in the event of a fall. Notably, wristbands and other components have lower energy capacities than smartwatches. Therefore, we used a smartwatch instead of other wearables as the primary solution for activity recognition in this study.

Human activity recognition (HAR) is a challenging research field focused on identifying the activity that an individual is engaging in, based on relevant activity data [9], [10]. Sensor-based HAR involves using data from a wearable sensor mounted on various locations of the body or placed in an individual's pocket. Such sensors may be embedded in smartphones, smart bands, and other wrist-worn devices such as smartwatches. Papers published in the last five years reflect the growing interest in HAR [11]–[14]. In the recent decade, most studies pertaining to HAR used traditional machine learning (ML) methods, including support vector machines, naïve Bayes, decision trees, k-nearest neighbor, and hidden Markov models, with ML algorithms. However, such machine learning techniques depend significantly on handcrafted shallow feature extraction, restricted by individual domain expertise [15]. Furthermore, traditional ML approaches segment and phase time-series data using different statistical formulas. Consequently, the temporal and spatial relationships are ignored during model training. Many recently published HAR techniques involved deep learning (DL) methods to address more complex HAR tasks. In addition, the availability of high-end graphics processing units (GPUs) enables the formulation of DL methods that can extract a larger number of upper-level characteristics from raw sensor data. Deep learning models have been used to realize automated feature extraction without using feature handcrafting to address the limitations of handcrafted feature extraction.

With their increasing use in recent years, deep learning techniques have been applied for HAR. In particular, convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) have demonstrated impressive data capture and fitting abilities in a variety of applications [16]. Convolutional operations allow CNNs to distinguish spatial features. Furthermore, CNNs are domain-independent and capable of generalization [17]. However, CNNs are computationally complex and require many training examples. Although CNNs collect the spatial characteristics of sensor data and provide the appropriate output for everyday human activities such as walking, jogging, seating, and standing [18], wearable sensor data must be considered to properly record actions that are too complicated to be captured by these devices. Temporal features are characteristics that are related to or vary with time. To recognize behavior from wearable sensors, time-series data are fed into recurrent neural networks (RNNs), which can detect temporal features [19]. LSTMs can mine long time-series dependencies, while CNNs excel at extracting local features [20]. Researchers have attempted to use LSTMs or CNNs to describe various forms of human behavior and achieved satisfactory results [21]. Although CNN and LSTM networks efficiently manage spatial and temporal information, their comprehension efficiency is restricted because their embeddings focus on specific data. CNNs with and without LSTMs yield comparable results, demonstrating that LSTMs do not effectively record temporal characteristics for HAR, owing to the lack of convolutional procedures [22].

In this work, ResNets [28] were used to facilitate learning processes in smartwatch-based HAR. In ResNets, a shortcut connection was introduced to effectively address the degradation problem of deep neural networks [29]. Specifically, a ResNet-based architecture for complex HAR was developed to classify complex physical activities such as eating, drinking, or smoking, based on a smartwatch sensor. Moreover, the squeeze-and-excitation mechanism [30] was incorporated in the proposed architecture to enhance the obtained information by feature recalibration. The proposed architecture represents the first ResNet-based framework for smartwatch-based activity identification.

WISDM-HARB, UT-Complex, and UT-Smoke datasets were used as benchmark datasets to validate the model performance. The proposed scheme achieved a high accuracy with increased durability under all datasets and experimentation settings. The key contributions of this study can be summarized as follows:

- Enhancement of a smartwatch HAR framework by using ResNets. This study represents the first attempt at establishing a ResNet-based structure as a generic framework for complex action recognition.
- A deep residual network, known as ResNet-SE, is designed for managing smartwatch sensor data and classifying complex human activities and tested on three publicly available datasets.
- Experiments are conducted to examine the influence of the window size on the model effectiveness in simple and complex tasks.

The remaining paper is organized as follows. The context theory for sensor-based HAR and DL methods is described in Section II. Section III describes the proposed methodology for complex HAR and deep learning method with an attention function. Section IV describes the experimental setup and presents the findings in terms of measurement metrics, which demonstrate the superior performance of the proposed approach due to the attention process. Section V discusses the

TABLE 1. Summary of HAR research related to SHAs and CHAs issues.

Work Ref.	Year	Simple Activity	Complex Activity
Shoaib et al. [23]	2016	walking; jogging; biking; writing; typing; sitting; standing	eating; drinking coffee; smoking; talking; ascending or descending stairs
Alo et al. [24]	2020	sitting; standing; walking; jogging; biking; ascending stairs,	descending stairs; eating; typing; writing; drinking; smoking, talking
Peng et al. [25]	2019	walking; running; sitting	having a meal; working; meeting; commuting; shopping; engaging in recreational activities; cleaning; exercising; sleeping
Liu et al. [26]	2016	sitting; standing; lying; ascending; descending; moving; walking; exercising; cycling; rowing; jumping	relaxing; drinking coffee; cleaning up; eating a sandwich; set-shot; jump-shot; lay-up; running dribbling; blocking; walk dribbling
Chen et al. [27]	2020	walking; sitting; standing	commuting; eating; cleaning

implications of the proposed system. Section VI presents the concluding remarks and highlights directions for future work.

II. THEORY BACKGROUND AND RELATED WORK

To perform classification and human action recognition, researchers have developed several learning-based approaches and conducted extensive HAR research [31]. Sensor-based activity recognition has been realized using deep learning to overcome the limitations of conventional machine learning methods in such applications [32]. According to existing studies [23]–[27], [33], human activities can be classified into two categories: simple human activities (SHAs) and complex human activities (CHAs).

Basic human activities, such as running, standing, or sitting, can be tracked using an accelerometer and identified [23]. Smoking, eating, and drinking are examples of complex human tasks that involve the use of the hands. Gyroscopes can distinguish between CHAs. This study labels stair climbing as a CHA [33] since it is impossible to distinguish the two actions with a single accelerometer. Strolling, running, sitting and standing are common human activities, as indicated by Alo *et al.* [24]. In contrast, shorter-duration actions such as smoking, chewing, taking ingesting medicine, dining, and writing correspond to complex behavioral patterns. Although this categorization does not fully represent the tasks performed in actual life, Peng *et al.* [25] adopted this scheme to categorize human interactions as simple or complex depending on whether they included repeated motions or single body postures. Tasks that need both simple and complex movements are more difficult to accomplish. Commonly, complex behaviors such as eating, working or shopping occur over a long period and include higher-level meanings. Consequently, these behaviors can provide a realistic depiction of a person's day-to-day activities. Liu *et al.* [26] found it challenging to simplify individual interactions. Two kinds of activities exist: activities that are temporally and theoretically connected, and activities that cannot be differentiated under software semantics. Moreover, a person can perform many tasks simultaneously instead of performing a single task at a time. Chen *et al.* [27] classified human endeavors as simple or complex. The accuracy of a single accelerometer is adequate to define the most basic human motions. Complex human actions are seldom as repeatable as simple activities, and they often involve many contemporaneous or overlapping motions that can only be detected by multimodal sensor data.

In this paper, we redefined these notions and established that simple human activities consist of repetitive motions devoid of hand gestures, whereas complex human activities consist of repetitive or nonrepetitive movements accompanied by hand gestures. SHA activities include walking, running, stair climbing, sitting, standing, and eating-related activities, whereas CHA activities include typing, clapping, and drinking. Table 1 summarizes the existing HAR research on SHAs and CHAs.

A. COMBINATION MODELS

According to recent experimental investigations, several deep learning architectures can be combined into a single model to achieve excellent HAR effectiveness. Xu *et al.* recommended the use of CNNs and gated recurrent units (GRUs) [34] to extract sequential temporal dependencies in complex action recognition [35]. Chen *et al.* [36] employed a 1D-CNN-LSTM network to extract deep features from lengthy acceleration sequences and an attention mechanism to integrate the handcrafted characteristics of heart rate variability data in a sleep-wake edge detector. Due to the unbalanced nature of the labeled data, an attention structure known as recurrent convolutional was presented as a semisupervised architecture by Chen *et al.* [37]. For action recognition, [38] applied a CNN over a small segment of window data and fed the retrieved features into an LSTM layer. In contrast to other models that apply the two-layer LSTM to the raw sensor data before introducing the 2D convolutional layer, the proposed framework did not utilize the features retrieved from CNN and instead used the feature representations from the LSTM over the HAR model [39].

In the recent year, a number of hybrid architectures, including InceptionTime [40], temporal transformer systems [41], and LSTM-FCNs [42], were developed to address specific time-series classification challenges. Actual data were used to categorize transportation-related activities using InceptionTime, which outperformed ResNet and CNNs in HAR research [41]. To address the HAR challenging task, Ronald *et al.* [43] used an Inception-ResNet model, which is an enhanced variant of the original framework.

III. PROPOSED METHODOLOGY

This part describes the procedure adopted to train a DL model and identify complex human activities through smartwatch built-in and wearable sensors based on the physical movement patterns. The proposed methodology for the ComplexHAR framework is shown in Figure 1. The model involves four phases: data acquisition, data preprocessing, training model, and model evaluation.

A. OVERVIEW OF THE FRAMEWORK

The proposed complex HAR framework employs sensor data from a wrist-worn wearable sensor to characterize the complex human behavior exhibited by smartwatch users. Figure 1 depicts the study design adopted to attain the study objectives. The framework operates through data acquisition, data preprocessing, feature representation, and model training/testing. Smartwatch sensor data from the WISDM-HARB dataset are collected, including SHAs and CHAs, for data acquisition. Sensor data are segregated with five sliding windows in the preprocessing data stage to produce data samples for the following step. A high-dimensional integrating space is used to generate feature representations for CNN/LSTM. Finally, attention-based HAR is used to enhance the recognition performance. The system is described in the following subsections.

B. DATASETS

The information of the three benchmark datasets is summarized in Table 2. Certain variations can be detected in the datasets in terms of complex human activities. In the case of the WISDM-HARB dataset, 51 people who wear smartwatches and engage in 18 activities provide the most accurate data. The UT-Smoke dataset consists of smartwatch sensor data recorded from 11 participants performing six activities. The UT-Complex consists of 13 activity sensor data of 10 subjects. In the UT-Complex dataset, wrist-worn sensor data are emulated using smartphone sensors at the wrist position of participants.

The WISDM-HARB dataset [44] contains smartphone sensing data collected from 51 subjects who were asked to perform 18 daily activities, including 5 simple activities and 13 complex activities. The sensor data were captured using a triaxial accelerometer and gyroscope sensors at a constant rate of 20 Hz while each subject performed the activities for 3 min.

The UT-Smoke was previously reported [45]. A smartwatch application was used to gather data from 11 individuals (two female and nine male participants with ages ranging from 20 to 45 y). In general, a wristwatch and a smartphone can be used to record the timestamp and triaxial accelerometer/gyroscope data. The sampling rate for all data was 50 Hz. Smoking while standing (SmokeSD), smoking while sitting (SmokeST), smoking while walking (SmokeW), smoking in a group (SmokeG), drinking while standing (DrinkSD), drinking while sitting (DrinkST), eating, standing, sitting, and walking (Walk). This dataset focuses on smoking in various forms and other actions that may be mistaken for smoking. Except for SmokeG and SmokeW, which were performed by 8 and 3 individuals, respectively, all activities were completed by all individuals.

The ‘‘Complex Human Activities using Smartphone and Smartwatch Sensors’’ accessible benchmark dataset is the third sensor dataset (UT-Complex dataset) [23]. Twente University, a pervasive system research center, made this data collection available to the public at the end of 2016. According to Table 2, the researchers collected data from 10 healthy volunteers who exhibited 13 human behaviors. To simulate a smartwatch, all ten volunteers were instructed to place two Samsung Galaxy S2 smartphones on their bodies, one phone in the right pocket of the trousers, and the other phone on the right wrist. To directly acquire sensor-based information, the participants were instructed to perform seven everyday tasks for three minutes. Seven of the ten individuals performed alternative difficult tasks, such as dining, typing, drawing, drinking, and conversing for 56 min. To examine the ability detect cigarette smoking, six volunteers were instructed to light up a cigarette. The researchers presented 30 min of data from each subject for each action to ensure an even distribution. Data from the accelerometer, gyroscope, and linear acceleration sensor were acquired at a sampling rate of 50 Hz.

C. DATA PREPROCESSING

Sensor-based HAR is initiated with the generation of data samples from raw sensor data. The brief periods are known as temporal windows, created by dividing the raw data into equisized blocks. The raw time series data acquired from wearable sensors are separated into temporal fragments prior to training a deep learning technique. The sliding technique is widely used and has been demonstrated to be effective with streaming data [46]. X, Y, and Z denote the three segments of a triaxial IMU sensor. The period is equal to the window size specified by Δt. D_t denotes the X, Y, and Z readings throughout the time interval $[t, \Delta t]$. The first sampling approach is referred to as a nonoverlapping temporal window (NOW), in which D_t and D_{t+1} are composed of many periods.

Because the temporal frames no longer overlap, the NOW approach allows only a small number of samples, i.e., $D_t \cap D_{t+1} = \emptyset$. The overlapping temporal window (OW) approach uses a fixed-size window to generate detailed samples from the sensor data series. The OW scheme, which has a 50% overlap proportion, is frequently used in sensor-based HAR research. Nevertheless, because D_t and D_{t+1} sections of the sensor readings are involved with an overlap fraction, this sampling is biased.

D. ResNet-SE NETWORK

The proposed ResNet-SE architecture is depicted in Figure 2. The fundamental architecture of the proposed deep learning model, aimed at to addressing the complex HAR problem, is composed of a convolutional block and five residual-SE blocks. One layer applies convolution to the sensor data, which are passed to another layer for batch normalization, a layer involving rectified linear units (ReLUs), and a layer for max-pooling in the convolutional component. The convolutional layer uses a variety of kernels, each of which

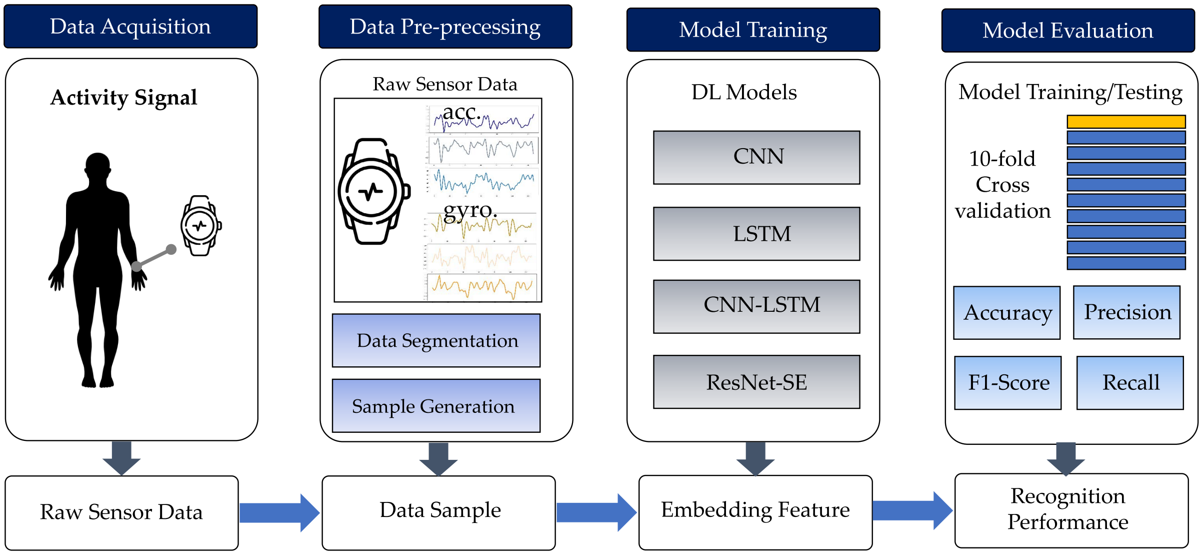

FIGURE 1. **Proposed ComplexHAR framework.**

TABLE 2. **Characteristics of the selected HAR datasets.**

Dataset	Sensors	Sensor Polling Rate	Number of Subjects	Number of Activities	Activities	
					Simple	Complex
UT-Smoke	Acc. Gyro.	50 Hz	11	6	standing sitting walking	smoking drinking eating
UT-Complex[a]	Acc. Gyro.	50 Hz	10	13	standing sitting walking jogging biking walking upstairs walking downstairs	typing writing drinking talking smoking eating
WISDM-HARB	Acc. Gyro.	20 Hz	51	18	sitting standing walking jogging stairs	typing brushing teeth eating soup eating chips eating pasta eating sandwiches drinking kicking playing dribbling writing clapping folding cloths

[a] A smartwatch sensor data is emulated by using a smartphone at the wrist position.

produces a feature map to collect a variety of distinct characteristics. The kernels, such as the input spectrum, are one dimensional. The primary goals of implementing the batch normalization layer are to stabilize and accelerate the training process. To enhance the model expressiveness, the ReLU layer is employed. To minimize the size of the feature map, a max-pooling layer is used to retain the most important features.

The result of the convolutional block is sent to the residual-SE block. By introducing a bypass connection within the residual-SE block, the deterioration issue can be efficiently addressed [47]. Layers of convolution, batch normalization, ReLU, and squeeze-and-excitation and a bypass connection are included in the residual-SE. The function of each component in the residual block is identical to that of the convolutional block, except for the bypass connectivity. The GAP is used to average each feature map, and the averaged entities are converted to a one-dimensional vector through a flattened layer in the proposed design. The outcome of the fully connected layer is processed using a softmax algorithm to assign a conditional probability to each group.

Figure 3 depicts the SE block, which is composed of two operations: a squeeze function, which aggregates the summarized information regarding each feature map, and an

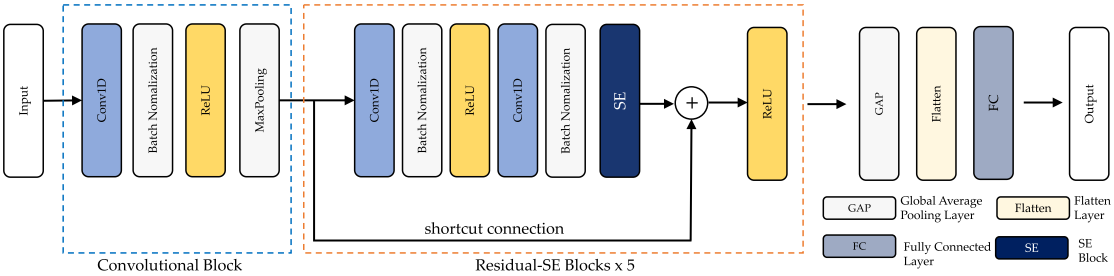

FIGURE 2. Proposed ResNet-SE architecture.

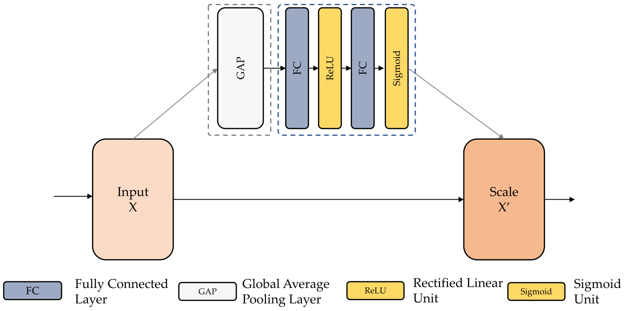

FIGURE 3. Residual-SE block.

excitation function, which modulates the relevance of each feature map according to its size. Utilizing global average pooling, the squeeze operation extracts only the most vital information from each channel, and the excitation operation calculates the interchannel dependencies by employing two fully connected layers and two nonlinear functions, namely, the ReLU and sigmoid functions, in a manner similar to the previous operation.

The values of hyperparameters in deep learning are employed to regulate the learning experience during model training. The proposed model made use of the following hyperparameters: (i) epochs; (ii) batch size; (iii) learning rate; (iv) optimization, and (v) loss function. To establish these hyperparameters, we specified the number of epochs to 200 and the batch size to 128. After 30 epochs, if no progress in the validation loss was seen, we implemented a call back of early stopping to bring the training process to an end. We began by setting the learning rate $\alpha = 0.001$. After six subsequent epochs, we adjusted this to 75% of its original value if the validation accuracy of the proposed model did not increase. To reduce error, the Adam optimizer [48] was employed with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10$. The optimizer utilized the categorical cross-entropy function to determine the error. The hyperparameter settings for the proposed ResNet-SE model are listed in Table 3.

TABLE 3. The summary of hyperparameters for the ResNet-SE network used in this work.

Stage	Hyperparameters		Values
Architecture	**Convolutional Block**		
	Convolution	Kernel Size	5
		Stride	1
		Filters	64
	Batch Normalization		-
	Activation		ReLU
	Max Pooling		2
	Residual-SE Block × 5		
	Convolution	Kernel Size	5
		Stride	1
		Filters	32
	Batch Normalization		-
	Activation		ReLU
	Convolution	Kernel Size	5
		Stride	1
		Filters	64
	Batch Normalization		-
	SE Module		-
	Global Average Pooling		-
	Flatten		-
	Dense		128
Training	Loss Function		Cross-entropy
	Optimizer		Adam
	Batch Size		64
	Number of Epochs		200

IV. EXPERIMENTAL RESULTS

This section explains the implementation of deep learning models such as CNN, LSTM, and CNN-LSTM and proposed ResNet-SE models for complex HAR. To evaluate the generalizability and effectiveness of the DL models, three publicly available datasets are employed. The sensor data for all the datasets are segmented using a fixed-length sliding window, which is a typical technique. The lengths of the sliding window are 5, 10, 20, 30, and 40 s, with an overlap proportion of 50%. Human activities pertaining to each dataset are classified into three categories: simple human activities (SHAs), complex human activities (CHAs), and all human activities (ALL), as shown in Table 2.

A. MODEL IMPLEMENTATION

We use the Google Colab Pro+ platform. The Tesla V100-SXM2-16GB graphics processor module is used to accelerate the training of the deep learning models. TensorFlow and CUDA are used to create the 1D-ResNet-SE and other fundamental deep learning techniques in the Python library. The following Python libraries are adopted:

TABLE 4. Performance of different DL models on UT-Smoke.

Window Sizes	Model	Recognition Performance					
		SHA		CHA		ALL	
		Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
5	CNN	96.97%	96.97%	95.38%	95.40%	95.41%	95.41%
	LSTM	99.96%	99.96%	95.12%	95.15%	94.60%	94.64%
	CNN-LSTM	99.96%	99.96%	90.93%	90.76%	91.19%	91.11%
	ResNet-SE	**99.98%**	**99.98%**	**98.41%**	**98.41%**	**98.84%**	**98.84%**
10	CNN	99.90%	99.89%	94.08%	94.09%	95.77%	95.76%
	LSTM	99.86%	99.86%	95.40%	95.44%	96.03%	96.06%
	CNN-LSTM	99.96%	99.96%	88.15%	87.95%	88.31%	88.14%
	ResNet-SE	**99.94%**	**99.94%**	**98.57%**	**98.57%**	**98.68%**	**98.68%**
20	CNN	99.92%	99.92%	94.48%	94.50%	94.71%	94.67%
	LSTM	99.76%	99.75%	97.45%	97.46%	97.24%	97.24%
	CNN-LSTM	**100.00%**	**100.00%**	89.29%	89.04%	89.77%	89.58%
	ResNet-SE	99.84%	99.84%	**98.68%**	**98.67%**	**99.13%**	**99.13%**
30	CNN	**100.00%**	**100.00%**	93.91%	93.87%	95.10%	95.03%
	LSTM	99.58%	99.58%	96.05%	96.07%	97.03%	97.03%
	CNN-LSTM	99.94%	99.94%	81.61%	79.50%	84.60%	83.57%
	ResNet-SE	**100.00%**	**100.00%**	**98.76%**	**98.76%**	**99.20%**	**99.20%**
40	CNN	99.92%	99.91%	94.12%	94.04%	94.50%	94.43%
	LSTM	99.60%	99.60%	97.21%	97.21%	95.59%	95.61%
	CNN-LSTM	99.84%	99.83%	85.13%	84.65%	85.77%	85.02%
	ResNet-SE	**100.00%**	**100.00%**	**99.31%**	**99.31%**	**99.23%**	**99.23%**

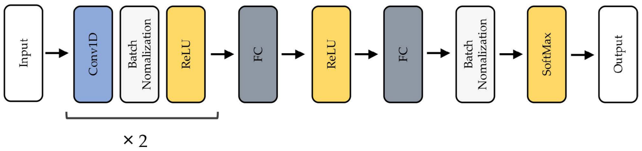

FIGURE 4. CNN architecture.

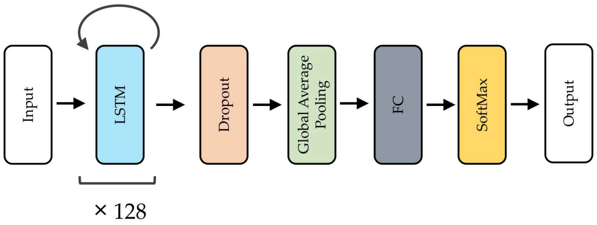

FIGURE 5. LSTM architecture.

- In analyzing the sensor data, Numpy and Pandas are used as data manipulation tools to retrieve, modify, and analyze the data.
- Matplotlib and Seaborn are used to plot and present the results of knowledge discovery and model evaluation.
- Scikit-learn (Sklearn) is used for sampling and data generation.
- To build and train deep neural networks, we use Keras, TensorFlow, and TensorBoard.

B. BASELINE DEEP LEARNING MODELS

This research evaluated our proposed ResNet-SE model against three benchmark deep learning models based on the CNN and LSTM architectures. The CNN model used in our study comprised two convolutional layers, a batch normalization layer, a ReLU activation layer, and two fully connected layers. Figure 4 illustrates a detailed description of the CNN architecture.

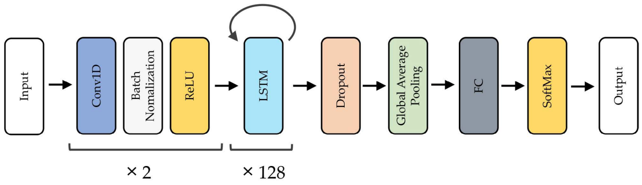

FIGURE 6. CNN-LSTM architecture.

The LSTM network was the investigation's second standard deep learning model. The structure of the LSTM was composed of a 128-cell LSTM layer, a dropout layer, an average pooling layer, and a fully connected layer. The LSTM, as shown in Figure 5, is a technique for resolving the vanishing gradient issue in long-term dependency learning. LSTM utilizes memory cells with three gates and parameters to describe long-range interdependence in temporal sequences. These gates determine when states are updated and when previously hidden conditions are forgotten, and therefore govern the memory cells' overall functionality.

The third baseline deep learning model is a CNN-LSTM hybrid model. The CNN-LSTM structure utilizes CNN layers to extract characteristics from the input data, while the LSTM segment handles sequence forecasting. The CNN-LSTM model can read subsequences acquired from the main sequence in the form of blocks by first extracting the key features from each block and then interpreting those characteristics using LSTM. The CNN-LSTM architecture employed in this study is shown in Figure 6.

TABLE 5. Performance of different DL models on UT-Complex.

Window Sizes	Model	Recognition Performance					
		SHA		CHA		ALL	
		Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
5	CNN	97.90%	97.89%	94.02%	94.02%	93.52%	93.52%
	LSTM	99.25%	99.24%	95.88%	**98.87%**	96.47%	96.48%
	CNN-LSTM	98.59%	**99.60%**	96.09%	96.10%	96.20%	96.19%
	ResNet-SE	**99.31%**	99.31%	**95.90%**	95.76%	**96.85%**	**96.84%**
10	CNN	97.82%	97.81%	95.97%	96.96%	93.31%	95.30%
	LSTM	98.29%	98.29%	94.90%	94.89%	95.81%	95.81%
	CNN-LSTM	98.93%	98.93%	**98.33%**	**98.33%**	**97.65%**	97.64%
	ResNet-SE	**99.76%**	**99.76%**	97.45%	97.42%	**97.65%**	**97.66%**
20	CNN	98.10%	98.09%	95.18%	95.17%	93.33%	93.32%
	LSTM	97.14%	97.15%	92.11%	92.11%	94.44%	94.43%
	CNN-LSTM	98.73%	98.73%	98.42%	98.41%	97.52%	97.51%
	ResNet-SE	**99.68%**	**99.68%**	**99.35%**	**99.35%**	**98.67%**	**98.67%**
30	CNN	96.07%	96.06%	91.64%	91.63%	91.98%	91.97%
	LSTM	95.71%	95.71%	89.84%	89.85%	90.44%	90.43%
	CNN-LSTM	98.57%	98.57%	**98.74%**	**98.75%**	97.69%	97.68%
	ResNet-SE	**99.64%**	**99.64%**	97.63%	97.65%	**98.01%**	**97.99%**
40	CNN	96.19%	96.20%	90.71%	90.70%	90.07%	90.06%
	LSTM	93.49%	93.48%	83.29%	83.28%	88.70%	88.69%
	CNN-LSTM	98.25%	98.25%	98.51%	98.51%	97.60%	97.60%
	ResNet-SE	**99.52%**	**99.52%**	**98.32%**	**98.32%**	**98.11%**	**98.10%**

TABLE 6. Performance of different DL models on WISDM-HARB.

Window Sizes	Model	Recognition Performance					
		SHA		CHA		ALL	
		Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
5	CNN	93.83%	93.82%	81.87%	81.88%	**92.33%**	**92.34%**
	LSTM	95.90%	95.91%	86.55%	86.56%	85.14%	85.14%
	CNN-LSTM	96.50%	96.49%	89.20%	89.19%	87.29%	87.28%
	ResNet-SE	**97.79%**	**97.79%**	**92.17%**	**92.13%**	91.80%	91.79%
10	CNN	92.71%	92.70%	79.58%	79.59%	77.97%	77.98%
	LSTM	94.89%	94.90%	88.26%	88.25%	86.97%	86.97%
	CNN-LSTM	96.55%	96.54%	91.66%	91.67%	90.33%	90.32%
	ResNet-SE	**97.55%**	**97.55%**	**95.18%**	**95.17%**	**95.07%**	**95.07%**
20	CNN	91.47%	91.47%	75.65%	75.66%	72.86%	72.85%
	LSTM	92.73%	92.74%	86.15%	86.14%	85.53%	85.53%
	CNN-LSTM	95.83%	95.83%	91.50%	91.49%	91.21%	91.22%
	ResNet-SE	**96.39%**	**96.39%**	**95.98%**	**95.99%**	**95.09%**	**95.08%**
30	CNN	90.87%	90.88%	72.57%	72.57%	68.97%	68.98%
	LSTM	89.16%	89.16%	83.10%	83.09%	83.41%	83.40%
	CNN-LSTM	**95.53%**	**95.53%**	90.56%	90.56%	90.98%	90.97%
	ResNet-SE	95.27%	95.26%	**95.88%**	**95.88%**	**94.89%**	**94.91%**
40	CNN	90.44%	90.44%	70.18%	70.18%	65.53%	65.52%
	LSTM	84.08%	84.09%	80.00%	79.99%	78.90%	78.89%
	CNN-LSTM	**95.73%**	**95.73%**	90.13%	90.13%	90.27%	90.26%
	ResNet-SE	95.15%	95.14%	**95.32%**	**95.31%**	**93.98%**	**93.98%**

C. ASSESSMENT ON PUBLIC DATASETS

The proposed framework is validated on three publicly available datasets to assess its performance. Tables 4, 5, and 6 show the classification results obtained when the model is evaluated over the UT-Smoke, UT-Complex, and WISDM-HARB datasets, respectively. The recognition performance is assessed separately for each category (SHA, CHA, and ALL) by using a five-fold cross-validation protocol with accuracy and F1-score metrics.

1) EXPERIMENTAL RESULTS ON THE UT-SMOKE DATASET

Various DL models, such as CNN, LSTM, and the proposed ResNet-SE model, are evaluated in the first experiment. Various segmentation sizes of 5, 10, 20, 30, and 40 s are set to train the deep learning models over the UT-Smoke dataset. The dataset consists of 3 simple activities (sitting, standing, walking) and 3 complex activities (smoking, drinking, eating). As indicated in Table 4, the proposed ResNet-SE outperforms the other DL models in terms of the accuracy (100.00% at window sizes of 30 and 40 s).

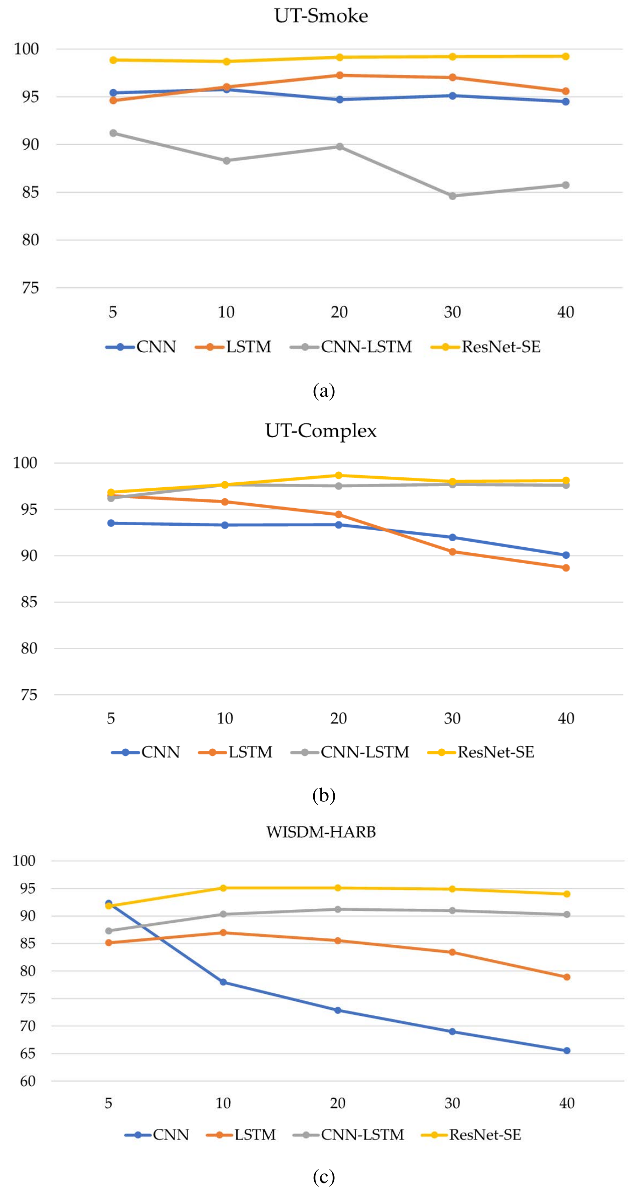

FIGURE 7. Effect of window sizes on (a) UT-Smoke, (b) UT-Complex and (c) WISDM-HARB.

2) EXPERIMENTAL RESULTS ON THE UT-COMPLEX DATASET

This experiment is based on the UT-Complex dataset, which consists of five simple actions (standing, sitting, strolling, running, and riding) and six complex activities (typing, writing, drinking, talking, smoking, and eating). The segmentation sizes are the same as those in the first experiment to evaluate the identification effectiveness of DL models. As indicated in Table 5, the proposed ResNet-SE achieves the highest accuracies for all segmentation sizes of sensor data. The highest accuracy of 99.68% corresponds to the SHA category with a sliding window of 20 s. For the CHA category, the highest accuracy is 99.35% with a sliding window of 20 s.

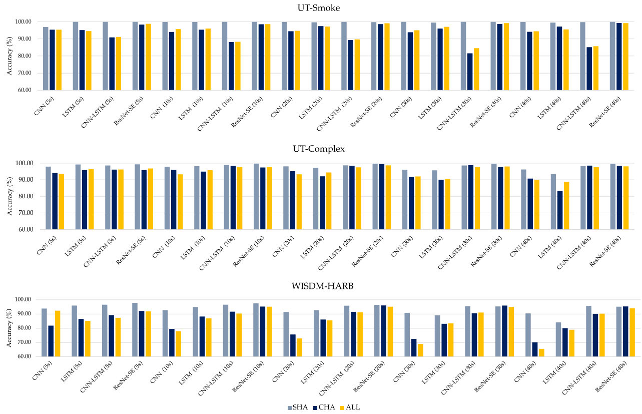

FIGURE 8. Comparison of results associated with different activities.

3) EXPERIMENTAL RESULTS ON THE WISDM-HARB DATASET

We use smartwatch sensor data from the WISDM-HARB dataset consisting of 5 simple activities and 13 complex activities, as shown in Table 6. Compared with SHA recognition, in CHA recognition, all baseline DL models (CNN, LSTM, and CNN-LSTM) and the proposed ResNet-SE model need sensor data segmented with larger window sizes. The highest accuracy of 97.79% corresponds to the ResNet-SE model with a window size of 5 s. In terms of CHA recognition, the proposed model outperforms other benchmark DL models with an accuracy of 95.98% for a size of 30 s.

V. DISCUSSION

This part discusses the findings presented in Section IV based on the experimental data.

A. *EFFECTS OF WINDOWS SIZES*

Several approaches, such as machine learning and DL windowing, are frequently employed to divide data in sensor-based HAR systems. A smaller window size corresponds to more efficient computing and smaller resource and energy consumption. Larger data windows are necessary for detecting more complex actions [49]. A window of 5 s is adequate to recognize simple tasks [23] such as running, standing, sitting, ascending and descending stairs, and strolling. Notably, an extremely small window size cannot discern the characteristics of complex activities such as typing or writing or sipping coffee or conversing. Changes in window size (5, 10, 20, 30, and 40 s) alter the training of DL models in a variety of contexts. Furthermore, as the size of the window increases, especially for more complex activities, an enhanced classification performance can likely be achieved, as shown in Figure 7.

B. *EFFECTS OF ACTIVITY TYPES*

In the experiments, we examine the influence of various types of tasks on the identification effectiveness. We select the three datasets because they feature two types of activities: basic

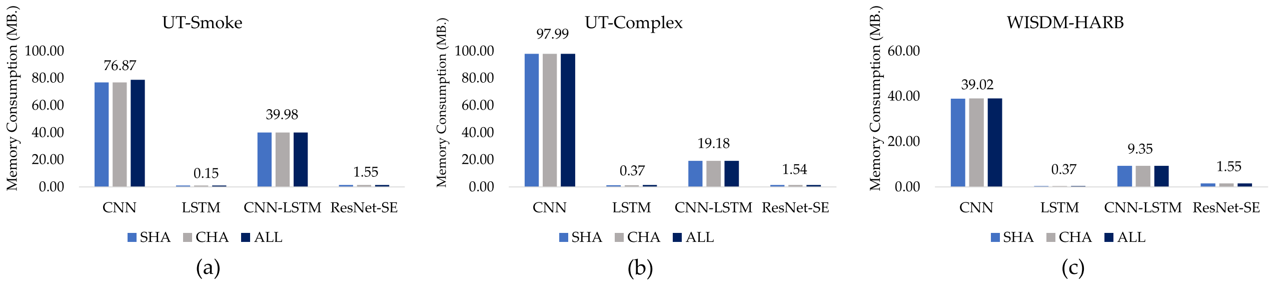

FIGURE 9. Memory consumption in megabytes of deep learning models used in this work, (a) UT-Smoke, (b) UT-Complex, and (c) WISDM-HARB.

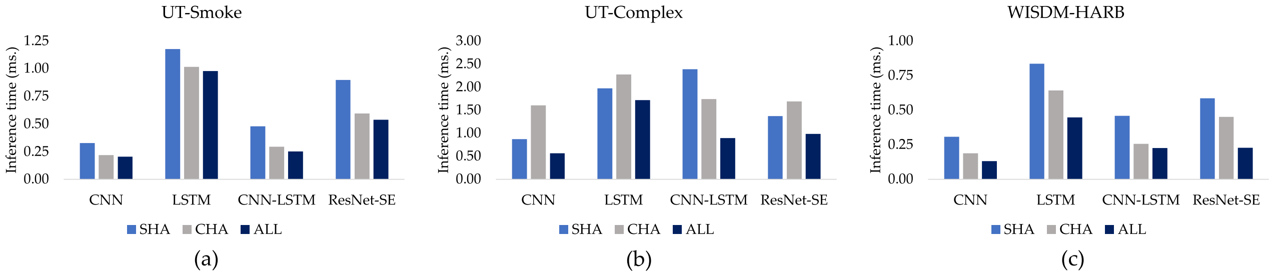

FIGURE 10. Mean prediction time in milliseconds of deep learning models used in this work, (a) UT-Smoke, (b) UT-Complex, and (c) WISDM-HARB.

activities and complex activities. Tables 4, 5, and 6 indicate that when trained on data of complex activities obtained from smartwatch sensors, all the evaluated deep neural networks attain the highest average accuracy, as shown in Figure 8. The proposed ResNet-SE model exhibits a high performance over the UT-Smoke, UT-Complex, and WISDM-HARB datasets, with average accuracies of 99.85%, 97.73%, and 94.9%, respectively.

C. COMPLEXITY ANALYSIS

We conducted a complexity study of the proposed deep learning model, including the baseline models, using the analysis technique for HAR stated in [50]. We defined model complexity in terms of memory consumption, mean prediction time, and the number of trainable parameters. This study validated all models against the same benchmark datasets (WISDM-HARB, UT-Smoke, and UT-Complex).

1) MEMORY CONSUMPTION

With today's smartwatches, memory consumption is less of a concern. For instance, the Apple Watch Series 7 and later models have 1 GB of RAM, whereas the Samsung Galaxy Watch 3 has 1.5 GB of RAM. As a result, establishing a smartphone application or a smartwatch application to implement HAR machine learning should be straightforward. After all, if we want to construct our wearable gadget, memory usage may be more vital if we reduce the size of the hardware or extend the power consumption. To track and compare memory use in this study, the deep learning models are deployed on an iPhone XR using the Tensorflow Lite framework, as recommended in [50]. The memory usage of each model could be calculated through an Xcode debug session. We tracked memory consumption and obtained the results indicated in Figure 9.

According to the comparative findings in Figure 9, the CNNs used the most memory while working with the three datasets. The LSTMs used the least memory, with values less than 1 MB. Using the proposed ResNet-SE model, each dataset requires approximately 1.55 MB.

2) PREDICTION TIME

We will continue to compare complexity to mean prediction time for efficiency consideration. To obtain the mean forecast, a series of samples from the testing data are input into the Tensorflow Lite networks, and the mean prediction time is then averaged.

Figure 10 demonstrates the experiment results with the mean prediction time in milliseconds to process one window of the deep learning models conducted on the three datasets (UT-Smoke, UT-Complex, and WISDM-HARB). The LSTM took the longest to arrive at a prediction with 0.98-1.74 ms. and 0.45-0.83 ms. for UT-Smoke and WISDM-HARB. When training LSTMs, it was seen that they take much more time for training than CNN-based models, including the proposed ResNet-SE. Convolutions can be accomplished in parallel. For computing the result of one kernel, only a few neighboring values are required. However, for an LSTM, much of the work needs to occur sequentially, as outcomes rely on the previous result. When considering the proposed ResNet-SE model, the mean prediction times were 0.54-0.90 ms., 0.98-1.69 ms., and 0.23-0.58 ms., for UT-Smoke, UT-Complex, and WISDM-HARB, respectively.

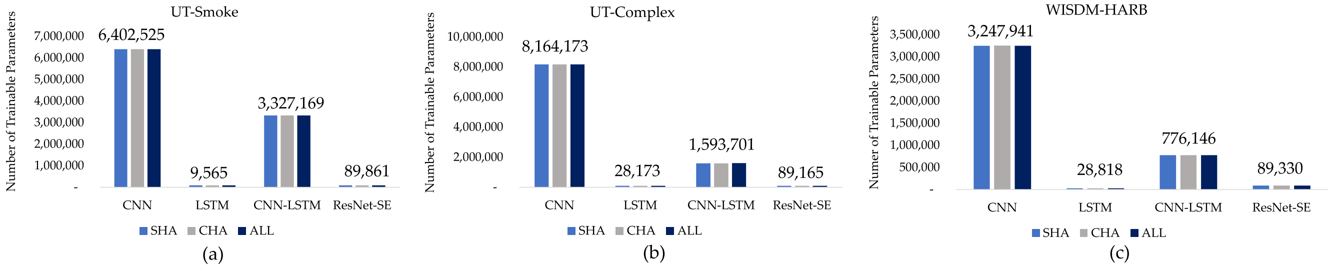

FIGURE 11. **Number of trainable parameters of deep learning models used in this work, (a) UT-Smoke, (b) UT-Complex, and (c) WISDM-HARB.**

3) TRAINABLE PARAMETERS

Considering memory consumption and mean prediction time, we can consider our third model complexity statistic, the number of trainable parameters. This technique is a statistic frequently employed with deep neural networks, where each weight learned during model training defines one such trainable parameter.

Figure 11 demonstrates the results of the trainable parameters of the deep learning models (CNN, LSTM, CNN-LSTM, and the proposed ResNet-SE) used in this work. The values can be obtained from the model summary of varied experimental scenarios (SHA, CHA, and ALL) on different standard HAR datasets (UT-Smoke, UT-Complex, and WISDM-HARB). The outcomes aligned with what we would intuitively anticipate when thinking of the complexity difference of these models. For the baseline deep learning models, the least complicated model according to the performed experiments is the LSTM with 9,565 parameters, 28,173 parameters, and 28,818 parameters for UT-Smoke, UT-Complex, and WISDM-HARB, respectively. In contrast to LSTM, CNN is the most complicated model with the most significant numbers of trainable parameters on the three datasets. Considering the proposed ResNet-SE model, the trainable parameters of this model are 89,861 parameters, 89,165 parameters, and 89,330 parameters for UT-Smoke, UT-Complex, and WISDM-HARB, respectively. However, the number of parameters of the ResNet-SE is lower than for the CNN and CNN-LSTM.

VI. CONCLUSION AND FUTURE WORKS

This study presents a sensor-based HAR model to recognize complex human activities by using smartwatch sensor data. The recognition performance of various DL models and the proposed ResNet-SE model is evaluated considering three benchmark datasets (UT-Smoke, UT-Complex, and WISDM-HARB) involving data of complex human activities recorded from smartwatch sensors. According to the findings, ResNet-SE outperforms the other DL models (CNN, LSTM, and CNN-LSTM) in all trials, independent of the number of layers in the network. The proposed ResNet-SE adopts channel attention through squeeze-and-excitation modules and shortcut connections to enhance the recognition performance in complex HAR tasks.

Future work will be aimed at extending and enhancing the model by optimizing the hyperparameters to decrease the model size and computation time. Furthermore, we plan to add spatial attention and channel attention mechanisms to the CNN networks to increase the identification accuracy.

REFERENCES

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[2] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *IEEE Access*, vol. 8, pp. 83791–83820, 2020.

[3] N. Rashid, M. Dautta, P. Tseng, and M. A. Al Faruque, "HEAR: Fog-enabled energy-aware online human eating activity recognition," *IEEE Internet Things J.*, vol. 8, no. 2, pp. 860–868, Jan. 2021.

[4] V. Senyurek, M. Imtiaz, P. Belsare, S. Tiffany, and E. Sazonov, "Electromyogram in cigarette smoking activity recognition," *Signals*, vol. 2, no. 1, pp. 87–97, Feb. 2021.

[5] S. Balli, E. A. Sağbaş, and M. Peker, "Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm," *Meas. Control*, vol. 52, nos. 1–2, pp. 37–45, Jan. 2019.

[6] P. Tarafdar and I. Bose, "Recognition of human activities for wellness management using a smartphone and a smartwatch: A boosting approach," *Decis. Support Syst.*, vol. 140, Jan. 2021, Art. no. 113426.

[7] Z. Wang, Z. Yang, and T. Dong, "A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time," *Sensors*, vol. 17, no. 2, p. 341, Feb. 2017.

[8] T. G. Stavropoulos, A. Papastergiou, L. Mpaltadoros, S. Nikolopoulos, and I. Kompatsiaris, "IoT wearable sensors and devices in elderly care: A literature review," *Sensors*, vol. 20, no. 10, p. 2826, May 2020.

[9] S. O. Slim, A. Atia, M. Elfattah, and M.-S. M. Mostafa, "Survey on human activity recognition based on acceleration data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 3, pp. 84–98, 2019.

[10] S. Mekruksavanich and A. Jitpattanakul, "LSTM networks using smartphone data for sensor-based human activity recognition in smart homes," *Sensors*, vol. 21, no. 5, p. 1636, Feb. 2021.

[11] F. Rustam, A. A. Reshi, I. Ashraf, A. Mehmood, S. Ullah, D. M. Khan, and G. S. Choi, "Sensor-based human activity recognition using deep stacked multilayered perceptron model," *IEEE Access*, vol. 8, pp. 218898–218910, 2020.

[12] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognit.*, vol. 108, Dec. 2020, Art. no. 107561.

[13] O. Nafea, W. Abdul, G. Muhammad, and M. Alsulaiman, "Sensor-based human activity recognition with spatio-temporal deep learning," *Sensors*, vol. 21, no. 6, p. 2141, Mar. 2021.

[14] M. A. R. Ahad, A. D. Antar, and M. Ahmed, "Sensor-based human activity recognition: Challenges ahead," in *IoT Sensor-Based Activity Recognition*. Cham, Switzerland: Springer, 2021, pp. 175–189.

[15] A. Sargano, P. Angelov, and Z. Habib, "A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition," *Appl. Sci.*, vol. 7, no. 1, p. 110, Jan. 2017.

[16] S. Mekruksavanich and A. Jitpattanakul, "Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models," *Electronics*, vol. 10, no. 3, p. 308, Jan. 2021.

[17] G. Zhang, G. Liang, F. Su, F. Qu, and J.-Y. Wang, "Cross-domain attribute representation based on convolutional neural network," May 2018, *arXiv:1805.07295*.

[18] F. Li, K. Shirahama, M. A. Nisar, L. Köping, and M. Grzegorzek, "Comparison of feature learning methods for human activity recognition using wearable sensors," *Sensors*, vol. 18, no. 2, p. 679, 2018.

[19] D. Singh, E. Merdivan, I. Psychoula, J. Kropf, S. Hanke, M. Geist, and A. Holzinger, "Human activity recognition using recurrent neural networks," in *Machine Learning and Knowledge Extraction*. Cham, Switzerland: Springer, 2017, pp. 267–274.

[20] S. Hochreiter and J. J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] S. Mekruksavanich, A. Jitpattanakul, P. Youplao, and P. Yupapin, "Enhanced hand-oriented activity recognition based on smartwatch sensor data using LSTMs," *Symmetry*, vol. 12, no. 9, p. 1570, Sep. 2020.

[22] I. Klein, "Smartphone location recognition: A deep learning-based approach," *Sensors*, vol. 20, no. 1, p. 214, Dec. 2019.

[23] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, "Complex human activity recognition using smartphone and wrist-worn motion sensors," *Sensors*, vol. 16, no. 4, p. 426, Mar. 2016.

[24] U. R. Alo, H. F. Nweke, Y. W. Teh, and G. Murtaza, "Smartphone motion sensor-based complex human activity identification using deep stacked autoencoder algorithm for enhanced smart healthcare system," *Sensors*, vol. 20, no. 21, p. 6300, Nov. 2020.

[25] L. Peng, L. Chen, Z. Ye, and Y. Zhang, "AROMA: A deep multi-task learning based simple and complex human activity recognition method using wearable sensors," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 1–16, Jul. 2018.

[26] L. Liu, Y. Peng, M. Liu, and Z. Huang, "Sensor-based human activity recognition system with a multilayered model using time series shapelets," *Knowl.-Based Syst.*, vol. 90, pp. 138–152, Dec. 2015.

[27] L. Chen, X. Liu, L. Peng, and M. Wu, "Deep learning based multimodal complex human activity recognition using wearable devices," *Appl. Intell.*, vol. 51, pp. 4029–4042, Jun. 2021.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[29] R. Monti, S. Tootoonian, and R. Cao, "Avoiding degradation in deep feed-forward networks by phasing out skip-connections," in *Proc. 27th Int. Conf. Artif. Neural Netw.*, Rhodes, Greece, Oct. 2018, pp. 447–456.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[31] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, Sep. 2018.

[32] Y. Chen, B. Zheng, Z. Zhang, Q. Wang, C. Shen, and Q. Zhang, "Deep learning on mobile and embedded devices: State-of-the-art, challenges, and future directions," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–37, Jul. 2021.

[33] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and complex activity recognition through smart phones," in *Proc. 8th Int. Conf. Intell. Environ.*, 2012, pp. 214–221.

[34] K. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.

[35] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan, "InnoHAR: A deep neural network for complex human activity recognition," *IEEE Access*, vol. 7, pp. 9893–9902, 2019.

[36] Z. Chen, M. Wu, W. Cui, C. Liu, and X. Li, "An attention based CNN-LSTM approach for sleep-wake detection with heterogeneous sensors," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 9, pp. 3270–3277, Sep. 2021.

[37] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semisupervised recurrent convolutional attention model for human activity recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1747–1756, May 2020.

[38] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Feb. 2020, pp. 362–366.

[39] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020.

[40] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. Schmidt, J. Weber, G. Webb, L. Idoumghar, P.-A. Müller, and F. Petitjean, "Inceptiontime: Finding AlexNet for time series classification," *Data Mining Knowl. Discovery*, vol. 34, no. 6, pp. 1936–1962, 2020.

[41] C. Naseeb and B. A. Saeedi, "Activity recognition for locomotion and transportation dataset using deep learning," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., ACM Int. Symp. Wearable Comput.* New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 329–334.

[42] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate LSTM-FCNs for time series classification," *Neural Netw.*, vol. 116, pp. 237–245, Aug. 2019.

[43] M. Ronald, A. Poulose, and D. S. Han, "ISPLInception: An inception-ResNet deep learning architecture for human activity recognition," *IEEE Access*, vol. 9, pp. 68985–69001, 2021.

[44] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133190–133202, 2019.

[45] M. Shoaib, H. Scholten, P. J. M. Havinga, and O. D. Incel, "A hierarchical lazy smoking detection algorithm using smartwatch sensors," in *Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2016, pp. 1–6.

[46] C. J. Chu, "Time series segmentation: A sliding window approach," *Inf. Sci.*, vol. 85, nos. 1-3, pp. 147–173, 1995.

[47] A. Hernández and J. M. Amigó, "Attention mechanisms and their applications to complex systems," *Entropy*, vol. 23, no. 3, p. 283, Feb. 2021.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–15.

[49] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, Apr. 2014.

[50] S. Angerbauer, A. Palmanshofer, S. Selinger, and M. Kurz, "Comparing human activity recognition models based on complexity and resource usage," *Appl. Sci.*, vol. 11, no. 18, p. 8473, Sep. 2021.

SAKORN MEKRUKSAVANICH (Member, IEEE) received the B.Eng. degree in computer engineering from Chiang Mai University, in 1999, the M.S. degree in computer science from the King Mongkut's Institute of Technology Ladkrabang, in 2004, and the Ph.D. degree in computer engineering from Chulalongkorn University, in 2012.

He is currently a Faculty Member with the Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phayao, Thailand. His current research interests include deep learning, human activity recognition, neural network modeling, wearable sensors, and applying deep learning techniques in software engineering.

ANUCHIT JITPATTANAKUL received the B.Sc. degree in applied mathematics from the King Mongkut's Institute of Technology North Bangkok, Bangkok, Thailand, and the M.Sc. degree in computational science and the Ph.D. degree in computer engineering from Chulalongkorn University.

He joined the Intelligent and Nonlinear Dynamic Innovations (INDI) Research Center, KMUTNB. He is currently a Faculty Member with the Department of Mathematics, King Mongkut's University of Technology North Bangkok. His current research interests include deep learning approaches applied to human activity recognition, wearable sensors, and healthcare applications.

KANOKWAN SITTHITHAKERNGKIET received the Ph.D. degree in mathematics from Naresuan University, Thailand.

She is currently a Lecturer with the Department of Mathematics, King Mongkut's University of Technology North Bangkok (KMUTNB). Her research interests include fuzzy optimization, fuzzy regression, fuzzy nonlinear mappings, least squares method, optimization problems, and image processing.

PHICHAI YOUPLAO received the B.Eng. degree in electrical engineering from North Eastern University, Khon Kaen, in 1998, the M.Eng. degree in electrical engineering from the Mahanakorn University of Technology, Bangkok, Thailand, in 2005, and the D.Eng. degree in electrical engineering from the King Mongkut's Institute of Technology Ladkrabang, Bangkok, in 2013.

He is currently an Assistant Professor with the Department of Electrical Engineering, Faculty of Industry and Technology, Rajamangala University of Technology Isan Sakon Nakhon Campus, Sakon Nakhon, Thailand. His current research interests include nano-devices and circuits, microring resonator, optical interferometry, quantum cryptography, sensors, and machine learning.

PREECHA YUPAPIN received the Ph.D. degree in electrical engineering from the City, University of London, U.K., in 1993.

He is currently a Full Professor with the Department of Electrical Technology, Faculty of Industrial Technology, Institute of Vocational Education Northeastern Region 2, Sakon Nakhon, Thailand. His current research interests include nano-devices and circuits, microring resonator, soliton communication, optical motor, quantum technologies, quantum meditation, and deep and machine learning.