Deep Pyramidal Residual Network for Indoor-Outdoor Activity Recognition Based on Wearable Sensor

Sakorn Mekruksavanich[1], Narit Hnoohom[2] and Anuchit Jitpattanakul[3,4,*]

[1]Department of Computer Engineering, School of Information and Communication Technology University of Phayao, Phayao, 56000, Thailand
[2]Image Information and Intelligence Laboratory, Department of Computer Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom, 73170, Thailand
[3]Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand
[4]Intelligent and Nonlinear Dynamic Innovations Research Center, Science and Technology Research Institute King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand
*Corresponding Author: Anuchit Jitpattanakul. Email: anuchit.j@sci.kmutnb.ac.th

Abstract: Recognition of human activity is one of the most exciting aspects of time-series classification, with substantial practical and theoretical implications. Recent evidence indicates that activity recognition from wearable sensors is an effective technique for tracking elderly adults and children in indoor and outdoor environments. Consequently, researchers have demonstrated considerable passion for developing cutting-edge deep learning systems capable of exploiting unprocessed sensor data from wearable devices and generating practical decision assistance in many contexts. This study provides a deep learning-based approach for recognizing indoor and outdoor movement utilizing an enhanced deep pyramidal residual model called SenPyramidNet and motion information from wearable sensors (accelerometer and gyroscope). The suggested technique develops a residual unit based on a deep pyramidal residual network and introduces the concept of a pyramidal residual unit to increase detection capability. The proposed deep learning-based model was assessed using the publicly available 19Nonsens dataset, which gathered motion signals from various indoor and outdoor activities, including practicing various body parts. The experimental findings demonstrate that the proposed approach can efficiently reuse characteristics and has achieved an identification accuracy of 96.37% for indoor and 97.25% for outdoor activity. Moreover, comparison experiments demonstrate that the SenPyramidNet surpasses other cutting-edge deep learning models in terms of accuracy and F1-score. Furthermore, this study explores the influence of several wearable sensors on indoor and outdoor action recognition ability.

Keywords: Human activity recognition; deep learning; wearable sensors; indoor and outdoor activity; deep pyramidal residual network

1 Introduction

Human activity recognition (HAR) has emerged as an increasingly exciting issue considering the recent introduction of intelligent wearable technology, wireless connectivity, and techniques for machine learning. Various applications will take advantage of the results of the HAR study, including athletic activity tracking [1], construction worker evaluation [2], and intelligent home monitoring [3]. In addition, HAR could be a viable opportunity for virtual medical systems for elderly individuals [4]. For instance, the HAR technology can offer care organizations valuable details and intelligent services so that the elderly can stay home for the longest time possible in a safe and healthy environment. Depending on the sensing technology employed, HAR techniques can be separated into two types: vision sensor-based and wearable sensor-based HAR. Vision-based HAR received the focus of researchers in video and image computation to determine HAR's satisfactory performance. Nevertheless, this strategy is impracticable for ordinary life since it would raise privacy issues. The solution can only function in the region where the camera is installed. Most of the research on HAR uses wearable sensors. Although the wearable sensor-based HAR has no privacy concerns and is environmentally insensitive, wearing the sensors for an extended period will make our everyday lives harder and involve considerable living expenses. Wearable technologies are increasingly employed in our everyday lives presently. The inertial sensors integrated with smart wearables are a viable alternative to wearable sensor-based HAR, allowing us to forego extra sensing elements [5]. Using the sensors from wearable technologies for HAR is thus an interesting topic to investigate.

Deep neural networks have significantly advanced sensor-based HAR applications in recent years. The capability of one to extract meaningful and express characteristics in a hierarchy from low-level to high-level abstractions has been demonstrated. Deep neural networks minimize the heuristic parameters of typical hand-designed characteristics and scalability more effectively for complicated behavior-recognition challenges. Previous findings on deep learning (DL) approaches for sensor-based activity detection have discovered that DL techniques outperform methods based on manually designed characteristics for human activity recognition [6]. For sensor-based HAR, Convolutional Neural Networks (CNNs) are the most often used DL technique. Even though HAR has been intensively examined, an appropriate feature learning strategy has yet to be exhaustively studied. Compared to the feature learning mentioned previously, a complex and deeper model structure enhances the accuracy of sophisticated HAR systems. These classifiers autonomously extract features using CNN. In object problems, a CNN feature extractor is often referred to as the backbone because the model architecture of the feature extractor and the overall model construction are examined independently. This work utilizes a CNN-based feature extractor as the "backbone" and self-attention as the sub-systems based on the formulation described in Section 3.

The following is a summary of the significant contribution that makes our proposed methodology superior to state-of-the-art methods in categorizing indoor and outside actions:

- SenPyramidNet, a deep pyramidal residual network, is presented for indoor and outdoor activity identification using wearable sensors (e.g., accelerometer and gyroscope). This proposed model functions as a hybrid of plain and residual networks by employing zero-padded identity-mapping shortcut interconnections while enhancing the dimension of the feature map.
- The recommended SenPyramidNet is optimized for spatial-temporal signals in terms of structure and training characteristics.
- We evaluate our proposed model using a standard HAR dataset, including mobility signals from both indoor and outdoor activities. The experimental findings demonstrate that our approach surpasses the current state-of-the-art approaches.

The following sections of this study are organized into six parts: Section 2 describes HAR's relevant research and intellectual context. The Section 3 covers the proposed methodology, which consists of a sensor-based HAR framework and a deep pyramidal residual model. Section 4 then analyzes the performance of several DL methods, including the proposed approach, based on controlled experiments using a standard dataset. In Section 5, we explore in depth the influence of sensor signals on the HAR platform's efficacy. Section 6 concludes with a summary of our findings and potential future research.

2 Related Works

There are two standard methods of HAR analysis: vision sensor-based HAR and wearable sensor-based HAR. Several features and insights can be obtained from this study, including the captured image and signals, extracted feature descriptors, and methodologies used for dimensionality reduction and human subject modeling.

2.1 Vision Sensor-Based HAR

Vision-based HAR relies entirely on visual sensing applications, such as observation and super-vision cameras, images or video sequences, modeling, segmentation, identification, and monitoring. Experimental findings [7] demonstrate that researchers presented a technique for identifying substantially more resilient and prosperous activities than the base classifiers. Nonetheless, the preliminary limitation of this research was the effectiveness of uncommon classes, namely transition movement classes. Moreover, they planned to enhance efficiency by including this class imbalance concern in their classification algorithm. Yang et al. [8] founded a new prototype for recognizing human actions in depth-camera-recorded video sequences. They also studied the low-level polynomial derived from a localized hyperspace in close proximity. In addition, their suggested scheme is customizable because it could be utilized in convergence with the combined trajectory-matched depth sequence. Their suggested model experienced a thorough analysis and evaluation of five standard datasets. The experimental results show that their suggested technique performs much better than current strategies on these datasets. Unfortunately, their suggested approach required additional data and incorporating numerous elements from both color and depth points to provide more cutting-edge representations. Moreover, Sharif et al. [9] offered an approach with two crucial phases. Initially, diverse individuals were uncovered in the captured video frame by integrating a new invariant segmentation with expectation-maximization (EM) segmentation. Employing vector dimensions, they retrieved and integrated localized features from determined sequences. A new Euclidean distance and joint entropy were used to select the ideal characteristics from the augmented vector. The optimum feature descriptors for motion identification were adjusted to the classifiers. Unfortunately, this effort did not solve the occlusions. Incorporating saliency to enhance segmentation precision is a different idea. Patil et al. in [10] suggested an approach for distinguishing and identifying human interactions in everyday life. In addition, they investigated numerous human visual databases to identify and track many human subjects. The background subtraction approach was used to track several moving individuals. Human everyday life operations with histogram of oriented gradient (HOG) feature extraction and a support vector machine (SVM) classifier gain superior recognition results with fewer false positives. Ji et al. [11] introduced a novel technique for interactive movement detection based on integrating probabilities at various stages. Furthermore, they addressed current complications in connection categorization techniques, such as insufficient feature descriptors due to incorrect human body segmentation. Consequently, a fusion technique based on many stages was developed to address this concern. However, this method is inefficient for handling the inherent qualities of human

interaction; it is excellent for identifying deviant behaviors, including violent actions and unexpected occurrences. Wang et al. introduced a probabilistic-based graphical approach for recognizing human physical activity in [12]. In addition, the problem of segmenting and identifying ongoing activity was considered. Unfortunately, these approaches only function offline. Using skeletal joint angle patterns, İnce et al. [13] produced a biometric system method for developing human physical activity in a three-dimensional environment. In addition, this framework uses the RGB-depth camera, which seems suited for video monitoring and senior care establishments. Nevertheless, there are a few limitations associated with the concept. Inadequate skeletal identification initially leads to inaccurate angle calculations and incorrect categorization.
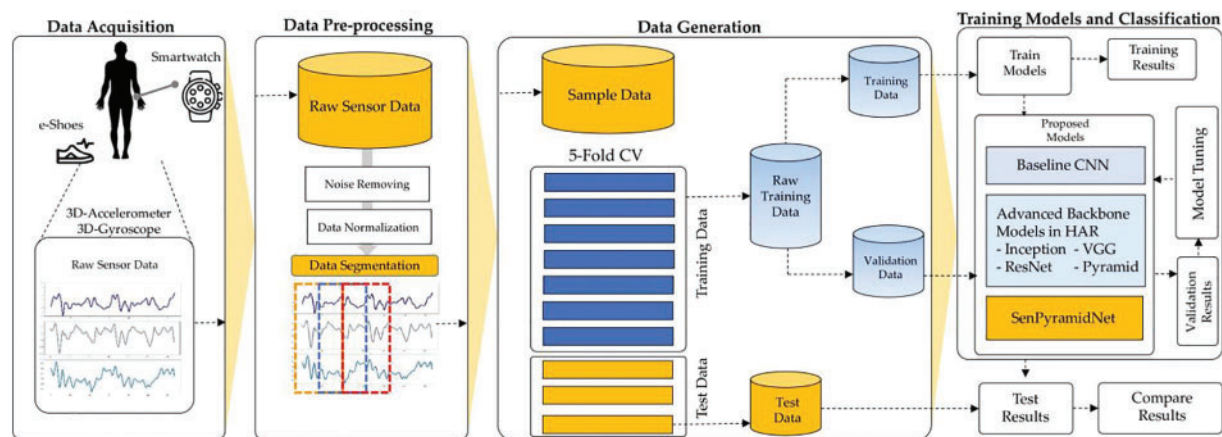
2.2 Wearable Sensor-Based HAR

Every aspect of our everyday life, from healthcare to convenience, has been changed by wearable inertial sensors. In this study, we evaluated IMU-based solutions due to the significant need for enhanced processing capacity and decreased space constraints. From the related literature, Irvine et al. [14] developed a homogeneous ensemble neural network technique for recognizing everyday actions in an indoor background. In addition, four standard models were created and fused through support function fusion. They also validated their suggested ensemble neural network technique framework by comparing the performance of human physical activity recognition (HPAR) achieved with two non-parametric conventional classifications. The robustness of the suggested ensemble approach was shown by the fact that the neural network ensemble methodology surpassed both conventional models. Nonetheless, the effort was limited since there needed to be a technique for selecting a meaningful subset of input attributes. Feng et al. presented an ensemble strategy for identifying HPAR [15] by combining a random forest approach with multiple wearable and inertial sensors. The random forest's increased predicting abilities contributed to a superior alternative for sensor-based wearable healthcare monitoring systems. Gupta et al. [16] introduced a successful bodily activity identification system based on a portable wearable accelerometer that could be applied to a real-world geriatric monitoring system. In addition, they integrated exceptional skills for identifying transitory actions. The suggested statistical characteristics collected extra details from the inertial signals inside the time-frame window. Signal correlation is extracted by evaluating other stimuli. Unfortunately, the introductory difficulty of this research is that just two individuals were utilized to collect data, which restricts the database's application in varied contexts. Abidine et al. [17] created a weighted SVM for recording human existence log events in an indoor setting. In addition, they handled other operational challenges associated with the HAR techniques, such as repetitive sequence features and group variations in the learning set. To overcome these issues, they provided a distinctive approach for detecting life log events indoors. In addition, the complete model was built on the fusions of many techniques, such as principal component analysis (PCA), SVM, and linear discriminant analysis (LDA). The PCA and LDA functions were used first to reduce the training set. Next, an SVM classifier was employed for each classification to optimize the classification performance on the imbalanced life log action database. De Cillis et al. [18] offered an innovative, pervasive approach for movement patterns utilizing a wearable inertial accelerometer sensor in a separate work. Their suggested model used a limited feature set and a decision tree classifier to identify four unique movement styles. They obtained characteristics from both static and dynamic collections of windows initially. The study's results demonstrated that accuracy was much higher while executing static activities than dynamic ones. The approach could be excellent for real-time medical operations due to its reduced computation overhead. Tian et al. [19] described a method for noticing bodily movements based on ensemble learning. Numerous characteristics were used to train three advanced classifiers and several SVMs,

possibly resulting in a system based on ensemble learning. In addition, an adaptive hybrid model retrieved numerous characteristics from an individual's body movements to boost their detection interpretation. Jung et al. [20] created a HAR-based system for recording life log everyday tasks and fall detection utilizing several wearable inertial sensors. Javed et al. [21] defined a state-of-the-art method for identifying physical movements based on sensory data collected from a two-axis smartphone accelerometer. In addition, the research considered the effectiveness and influence of the particular accelerometer axis in identifying human physical behaviors. Moreover, this approach includes multi-modal sensory data gathered from three body-worn sensors. This research illustrates that adding data from inertial sensors increases the HAR's accuracy. The system was evaluated through an exhaustive collection of cyclic, static, and random activities. Temporal and frequency domain variables were retrieved to achieve the most reasonable outcomes.

3 Proposed Methodology

This section details the process designed to directly construct a deep learning model and identify driving-related activities using smart glasses' built-in sensors. Fig. 1 depicts the proposed strategy for the smart glasses HAR configuration. It includes the five processes of data collection, preprocessing, training the learning model, and action recognition.

**Figure 1:** The proposed HAR methodology based on smart glasses sensors

3.1 Data Acquisition

The initial step of the HAR procedure is the collection of sensor data captured by wearables. This approach chose the 19NonSens dataset [22] as the public standard HAR dataset. This wearable sensor dataset for individual action detection is compiled from e-shoes, smart glasses, and a smartwatch. Twelve participants between the ages of 19 and 45 are requested to wear e-Shoes and a Samsung Gear-S2 wristwatch on their dominant hand (ten right-handed and two left-handed). The participants are given a list of 18 events—nine inside hobbies, like brushing and chopping, and nine outside sports, including kicking and sprinting. The participant might undertake any 18 tasks in the task list during pre-defined tasks. The dataset regarded all irrelevant actions as Null actions. The period of each exercise ranges between 3 and 10 min. Table 1 provides an overview of the particulars of each 19NonSens dataset.

**Table 1:** A list of activities in the 19NonSens dataset

No.	Activity	Environment	Movement	Duration (minutes)
1	Brushing	Indoor	Upper	78
2	Washing hand	Indoor	Upper	48
3	Slicing	Indoor	Upper	71
4	Peeling	Indoor	Upper	72
5	Walking upstairs	Indoor	Upper and lower	58
6	Walking downstairs	Indoor	Upper and lower	56
7	Mixing	Indoor	Upper and lower	52
8	Wiping	Indoor	Upper and lower	63
9	Sweeping floor	Indoor	Upper and lower	102
10	Turning shoulder	Outdoor	Upper	46
11	Turning wrist	Outdoor	Upper	51
12	Turning knee	Outdoor	Lower	53
13	Turning haunch	Outdoor	Lower	52
14	Turning ankle	Outdoor	Upper and lower	49
15	Walking	Outdoor	Upper and lower	64
16	Kicking	Outdoor	Upper and lower	77
17	Running	Outdoor	Upper and lower	61
18	Cycling	Outdoor	Lower	76

In the 19NonSens dataset, the accelerometer and gyroscope signals will be utilized as the scheme's input information. Both the accelerometer and gyroscope sensors of the Samsung Gear G2 are configured to a reading frequency of 50 Hz, similar to the sampling rate of the 3-axis wireless accelerometers, to facilitate synchronization. Several unprocessed data captured by the Samsung Gear G2 are depicted in Figs. 2 and 3. The wearable sensor used in the e-Shoe is a micro-electro-mechanical system (MEMS) accelerometer. The sampling frequency of the 3D-accelerometer data captured from the e-shoes is adjusted to 50 Hz. Several raw data captured by the e-Shoe are illustrated in Fig. 4.
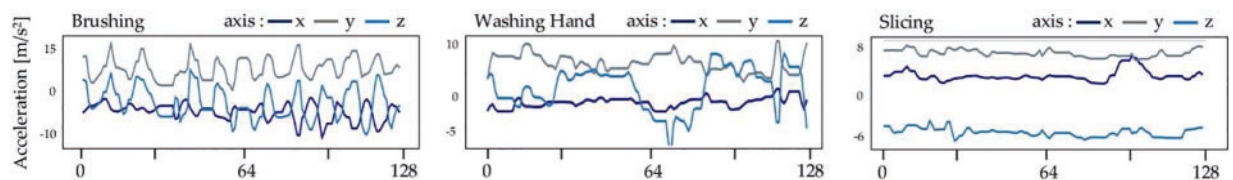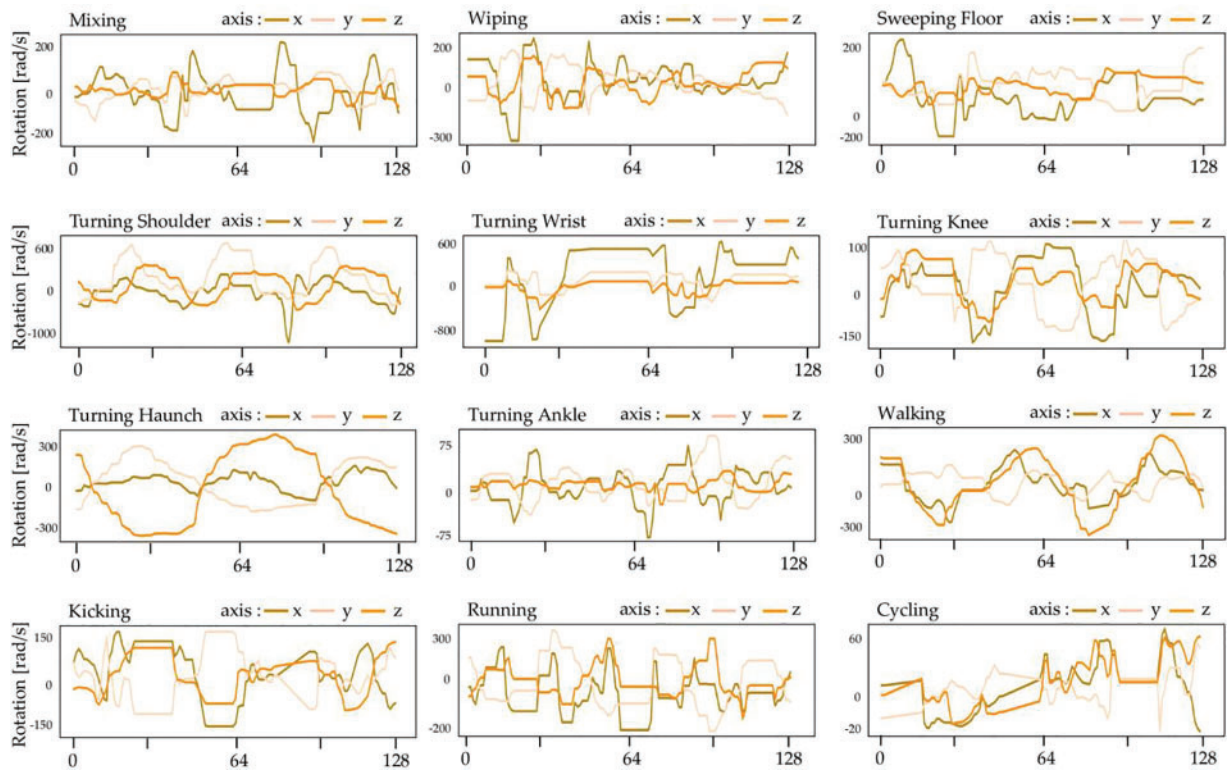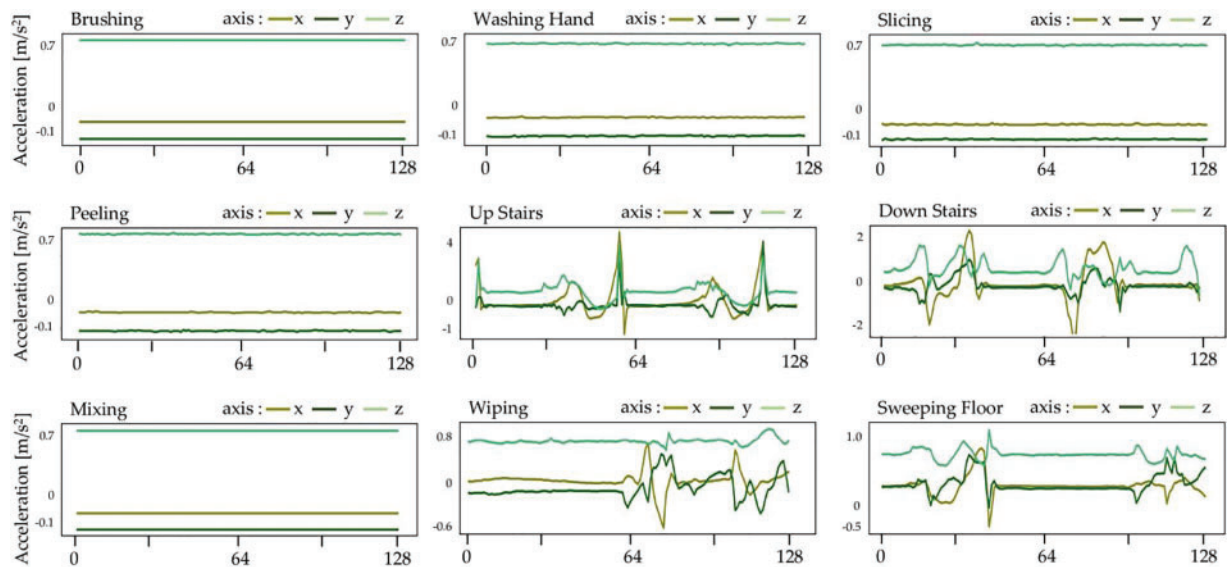

Figure 2: (Continued)
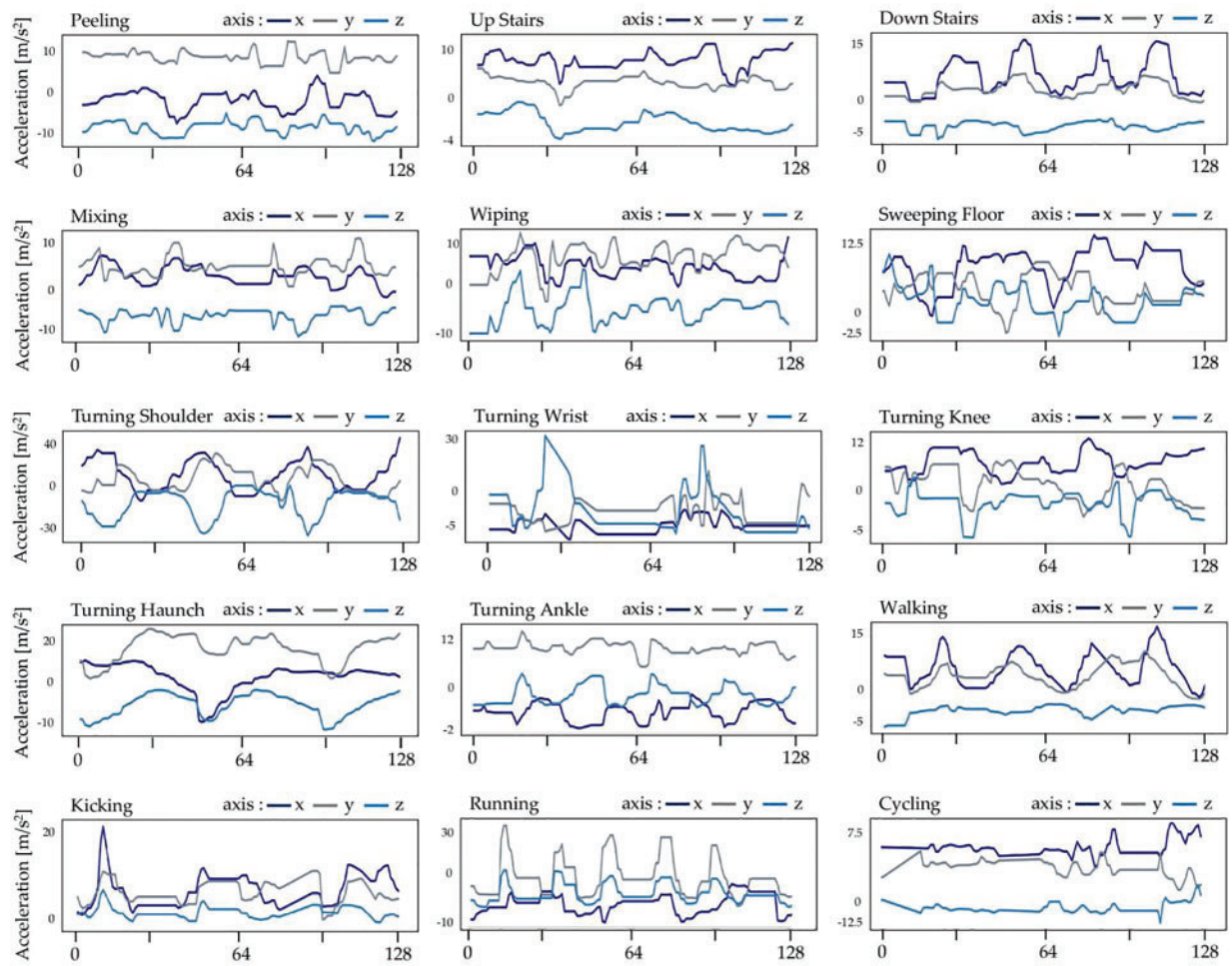
Figure 2: Some raw accelerometer data recorded from Samsung Gear G2 in the 19Nonsens dataset
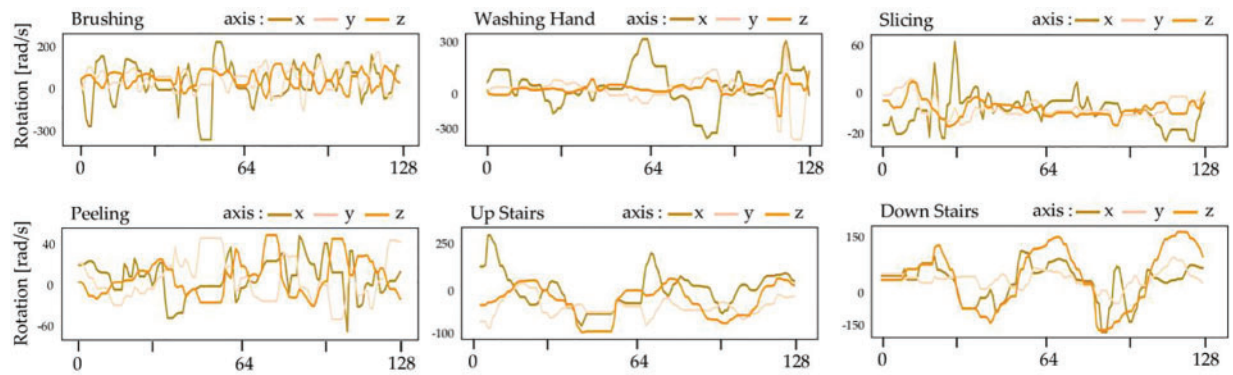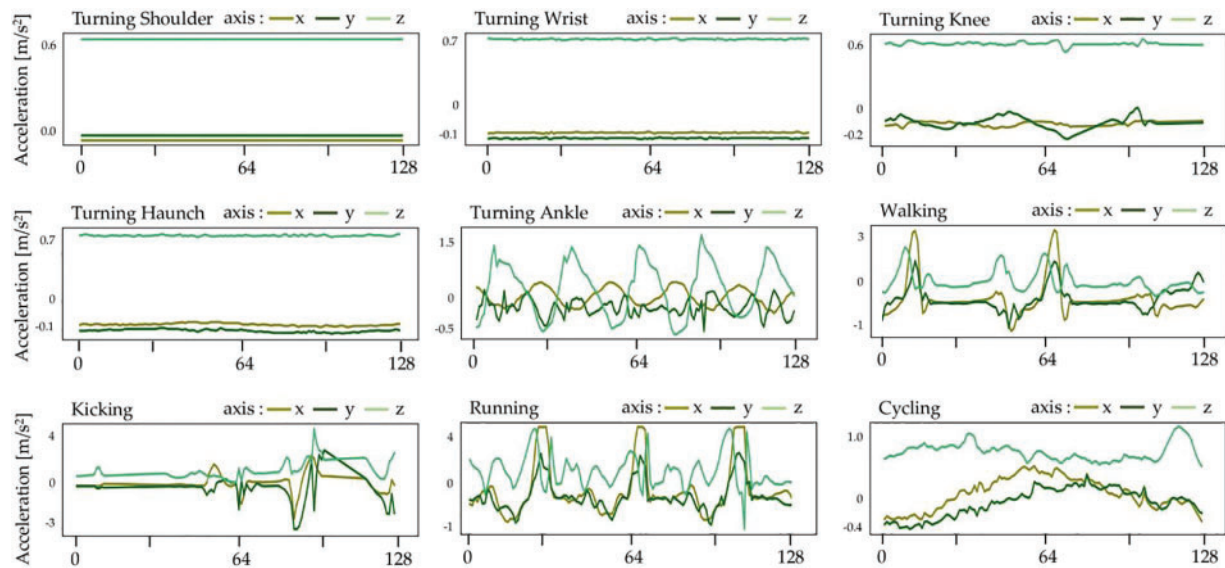

Figure 3: (Continued)

Figure 3: Some raw gyroscope data recorded from Samsung Gear G2 in the 19Nonsens dataset

Figure 4: (Continued)

Figure 4: Some raw accelerometer data recorded from e-Shoe in the 19Nonsens dataset

3.2 Pre-Processing Process

According to the person's active movements while data gathering, the raw data obtained by the wearable sensors include measurement noise and other unexpected interference. Relevant information inside a transmission is obscured by noise. Consequently, it was crucial to limit the impact of noise on the motion to gather crucial data for subsequent processing. The most frequently employed filtering methods are mean, low-pass, and Wavelet [23]. We deployed a third-order low-pass Butterworth filter with a 20 Hz cutoff frequency on the accelerometer and gyroscope sensors in all multiple dimensions to denoise the data. Considering that 99.9% of the energy is contained below 15 Hz, this rate is enough for capturing bodily motions.

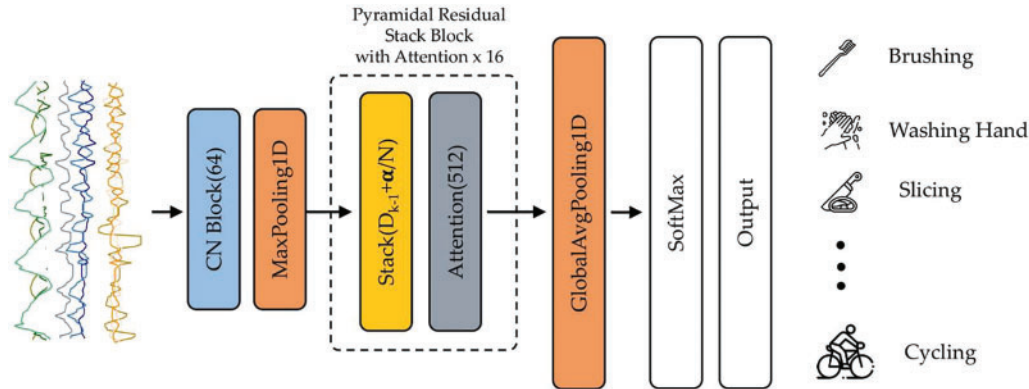### 3.3 The Proposed SenPyramidalNet

SenPyramidalNet, as seen in Fig. 5, is a deep residual network designed to effectively classify indoor and outdoor activities using wearable sensors. This proposed model is based on Han's 2017 [24] proposal for deep pyramidal residual networks (DPRN).

Based on the residual network, the DPRN optimizes the residual unit model. The primary point of the DPRN is to focus on the feature map dimension by steadily increasing it instead of drastically raising it at each residual unit by downsampling. Most deep convolutional neural network designs enhance feature map dimensions by a substantial margin as the scale of the feature map reduces and does not raise feature map dimensions unless they meet a layer with downsampling. To address this issue, DPRN attempts to progressively expand the feature map size, rather than increasing it at one of the residual units, and to divide the rising feature map load equally. This configuration results in a steady rise in the number of channels as a proportion of the layer's depth, akin to a pyramid whose base steadily expands towards its apex. In our investigations, the proposed SenPyramidNet selects the

additive PyramidNet model, which enhances the dimension of the feature map proportionally. The equation for enhancing the dimensions of a feature map is Eq. (1).

$$
D_k = \begin{cases} 64, & \text{if } k = 1 \\ \left\lfloor D_{k-1} + \dfrac{\alpha}{N} \right\rfloor, & \text{if } 2 \le k \le N+1 \end{cases} \tag{1}
$$

N represents the total number of residual units, and α represents a step factor for increasing dimensions.

Figure 5: The architecture of the proposed SenPyramidNet model

In addition, the network design functions as a combination of plain and residual networks by employing zero-padded identity-mapping shortcut connectivity when the feature map dimension is increased. Several configurations of ResNet shortcuts, including an identity-mapping shortcut, were examined. The identity-mapping shortcut is much more suitable than other alternatives. Due to the absence of parameters, an identity-mapping shortcut has a lesser likelihood of overfitting than other shortcuts, resulting in enhanced generalization performance. Furthermore, this could pass through the gradient strictly based on the uniqueness mapping, providing higher consistency throughout the training phase. In the case of SenPyramidNet, identity mapping alone cannot be employed as a shortcut since the dimension of the feature map varies across residual units. Moreover, the zero-padded shortcut does not result in the overfitting issue since there are no new parameters, and interestingly, it demonstrates a greater capacity for generalization than other shortcuts. Consequently, the suggested zero-padded identity-mapping shortcut depicted in Fig. 6 could deliver a mixed impact of the residual network and the plain network that is significantly enhanced.

As seen in Fig. 6, the proposed model's convolutional unit comprises a one-dimensional convolution layer (Conv1D) and a batch normalization (BN) layer. Each kernel produces a feature map, and One-dimensional kernels, for example, the input spectrum, were included. The use of BN stabilized and accelerated the training process. Each feature map was turned into a 1D vector employing a flattened layer after being averaged utilizing GAP. A softmax function was utilized to convert the completely integrated layer's output into probabilistic reasoning for each class. The network was trained to utilize the Adam optimizer, and the loss was calculated using the cross-entropy loss function, frequently employed in classification.
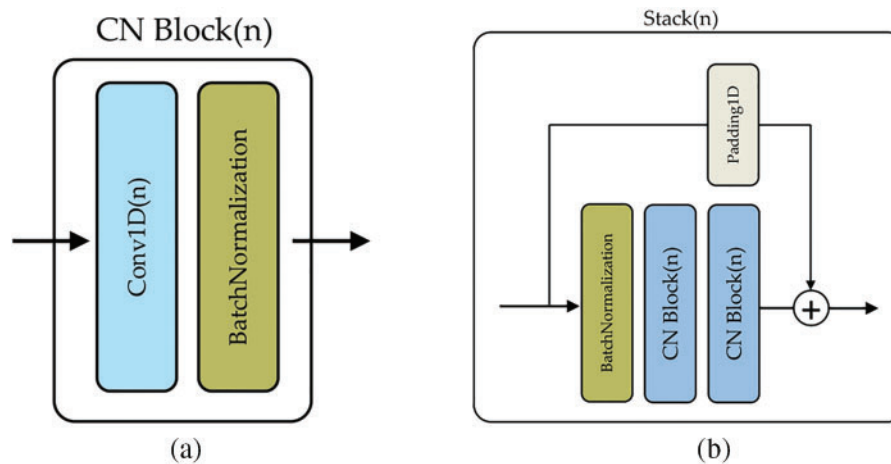
Figure 6: Schematic structure of (a) a convolutional block, and (b) a stack block in the proposed SenPyramidNet

4 Evaluations and Results

This section contains the outcomes of all experimental investigations undertaken to determine the most effective DL models for indoor and outdoor movement identification and the studies' results. Our evaluations were conducted using a standard dataset known as 19Nonsens that captured motion signals from a variety of activities both indoors and outdoors. To assess the proposed SenPyramidNet, we performed two studies based on 10-fold cross-validation using the following dataset activity data:

- Study I: We employed indoor motion data to validate the proposed model with state-of-the-art advanced models (Inception-ResNet, Inception, Xception, VGGs, and ResNets).
- Study II: The second study was accomplished utilizing motion signals of outdoor activities to evaluate the detection capability of state-of-the-art advanced models with our suggested SenPyramidNet model.

4.1 Environmental Configuration

Evaluations were conducted using the Google Colab Pro+ [25] framework. The results of accelerating DL model training with the Tesla V100-SXM2 with a 16GB graphics processor module were outstanding. The proposed SenPyramidNet and advanced DL models were developed using the Python library, the Tensorflow backend, and the CUDA graphics cards. The Python libraries that produced this investigation are mentioned below:

- The sensor data were operated employing Numpy and Pandas, which included reading, processing, and analyzing the data.
- Matplotlib and Seaborn were manipulated to plot and display the consequences of the data finding and model assessment procedures.
- Sklearn framework was utilized for sampling and data creation.
- TensorFlow, Keras, and TensorBoard were operated to create and train DL models.

4.2 Hyperparameter Setting

Hyperparameters are manipulated to handle the learning process in DL. The hyperparameters utilized in the proposed SenPyramidNet model are (1) the number of epochs, (2) batch size, (3) learning

rate α, (4) optimization, and (5) loss function. For setting the values of these hyperparameters, the number of epochs was set to 200, and the batch size was assigned to 128. If no improvement in the validation loss was uncovered after 20 epochs, we quit the training process with an early-stop callback. Initially, we arranged the learning rate $\alpha = 0.001$. Then, we updated it to 75% of the initial value if the validation accuracy of the proposed model remained the same after six consecutive epochs. Adam optimizer [26] is set with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1 \times 10^{-8}$ to minimize the error.

4.3 Experimental Results of Indoor Activity Recognition

The results of Study I used motion signals of indoor activities captured by an accelerometer integrated with an e-Shoe and an accelerometer and gyroscope of a smartwatch for training and evaluating models. Table 2 summarizes the findings of the SenPyramidnet and state-of-the-art models.

Table 2: Experimental studied results of DL models using indoor activity data

Model	Performance (Mean ($\pm$Std.))		
	Accuracy	Loss	F1-score
Inception-ResNet	92.17% ($\pm$0.89%)	0.38 ($\pm$0.04)	91.25% ($\pm$1.05%)
Inception	95.52% ($\pm$0.56%)	0.19 ($\pm$0.03)	94.99% ($\pm$0.60%)
Xception	93.79% ($\pm$0.59%)	0.31 ($\pm$0.03)	93.06% ($\pm$0.59%)
VGG11	76.57% ($\pm$13.41%)	0.88 ($\pm$0.32)	73.22% ($\pm$15.31%)
VGG13	90.89% ($\pm$2.28%)	0.47 ($\pm$0.06)	89.68% ($\pm$2.95%)
VGG16	85.08% ($\pm$5.56%)	0.88 ($\pm$0.54)	83.71% ($\pm$6.07%)
VGG19	81.48% ($\pm$16.75%)	0.65 ($\pm$0.50)	79.26% ($\pm$19.52%)
ResNet16	96.28% ($\pm$0.47%)	0.18 ($\pm$0.04)	96.13% ($\pm$0.50%)
ResNet18	95.69% ($\pm$0.87%)	0.21 ($\pm$0.09)	95.57% ($\pm$0.94%)
ResNet34	96.15% ($\pm$0.36%)	0.19 ($\pm$0.02)	96.03% ($\pm$0.38%)
SenPyramidNet	96.37% ($\pm$0.35%)	0.17 ($\pm$0.01)	96.08% ($\pm$0.39%)

Based on comparison findings in Table 2, the proposed SenPyramidNet model surpassed the other DL models in this investigation, achieving the most fantastic accuracy (96.33%) and F1-score (96.08%). Comparing the core structure of state-of-the-art models, we discover that ResNets have more accuracy than other models. This model utilizes the residual architecture while reinforcing the feature at each convolutional level.

4.4 Experimental Results of Outdoor Activity Recognition

Study II indicated that only outside motion data is accessible for training and testing DL models, as summarized in Table 3. In this experiment, the proposed SenPyramidNet model achieved the most incredible accuracy of 97.25% and the highest F1-score of 97.95%, outperforming other DL models.

Table 3: Experimental studied results of DL models using outdoor activity data

Model	Performance (Mean ± Std.)		
	Accuracy	Loss	F1-score
Inception-ResNet	94.56% (±0.90%)	0.26 (±0.05)	94.61% (±0.86%)
Inception	93.75% (±1.20%)	0.29 (±0.06)	93.89% (±0.94%)
Xception	95.04% (±1.16%)	0.30 (±0.06)	95.03% (±1.10%)
VGG11	88.58% (±1.12%)	0.80 (±0.17)	88.26% (±1.20%)
VGG13	89.16% (±1.36%)	0.68 (±0.12)	89.26% (±1.56%)
VGG16	86.46% (±1.76%)	0.69 (±0.10)	86.87% (±1.81%)
VGG19	86.34% (±1.88%)	0.61 (±0.10)	86.72% (±1.89%)
ResNet16	95.66% (±1.48%)	0.29 (±0.12)	95.72% (±1.12%)
ResNet18	96.49% (±0.67%)	0.32 (±0.08)	96.37% (±0.71%)
ResNet34	96.27% (±0.47%)	0.31 (±0.05)	96.10% (±0.35%)
SenPyramidNet	97.25% (±0.43%)	0.19 (±0.03)	97.95% (±0.47%)

5 Research Discussion

5.1 Effects of Different Wearable Sensors

To determine the influence of wearable sensors on the effectiveness of the proposed SenPyra-midNet models, we performed further investigations employing e-Shoe and smartwatch sensors independently. The comparison findings are summarized in Tables 4 and 5. These findings suggest that smartwatch sensors are more suitable than sensors incorporated in e-Shoes for training indoor and outdoor HAR models.

Table 4: Comparative results of the SenPyramidNet using different wearable sensors separately to recognize indoor activities

Wearable sensor	Performance (Mean ±Std.)		
	Accuracy	Loss	F1-score
e-Shoe	64.65% (±2.34%)	1.29 (±0.27)	67.82% (±2.75%)
Smartwatch	95.21% (±0.28%)	0.19 (±0.02)	94.52% (±0.32%)

Table 5: Comparative results of the SenPyramidNet using different wearable sensors separately to recognize outdoor activities

Wearable sensor	Performance (Mean ±Std.)		
	Accuracy	Loss	F1-score
e-Shoe	89.21% (±2.27%)	0.58 (±0.11)	86.15% (±3.21%)
Smartwatch	96.82% (±0.64%)	0.15 (±0.05)	97.35% (±0.57%)

To determine the causes of the divergent performances, we study the confusion matrices for each individual, as seen in Fig. 7. The suggested model obtains an F1-score of 67.82% using solely e-Shoe data. Most indoor tasks are hand-oriented, including brushing, peeling, and slicing. As demonstrated in Fig. 7, the movement signals collected by e-Shoe sensors cannot be distinguished. Therefore, the sensors of smartwatches are suited for indoor activity identification.
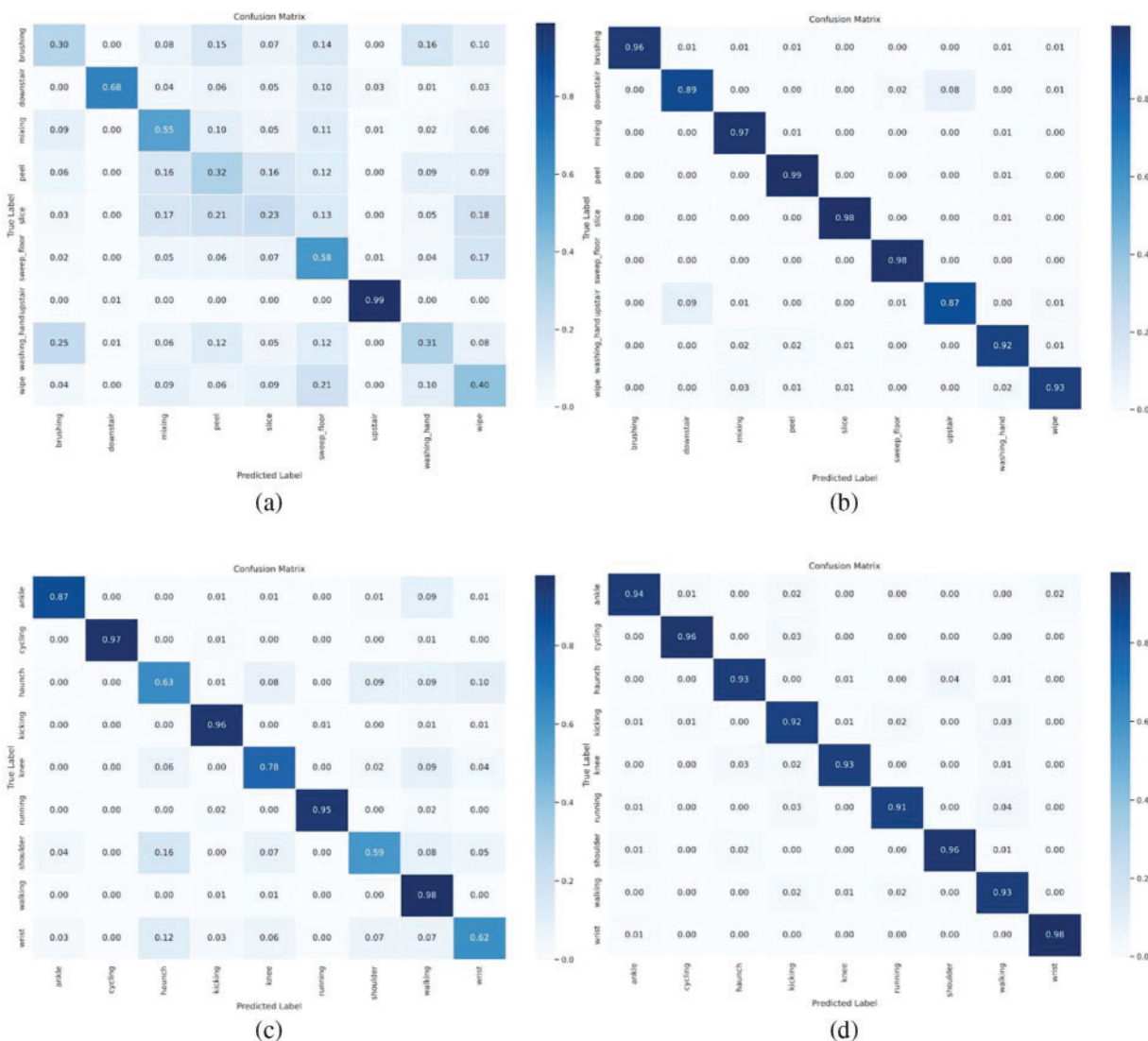
Figure 7: Comparison of the model performance between confusion matrices: (a) indoor HAR using e-Shoe sensors (b) indoor HAR using Smartwatch sensors (c) outdoor HAR using e-Shoe sensors (d) outdoor HAR using Smartwatch sensors

5.2 Effects of Attention Mechanism

Most learning-based solutions require the capability to learn an open-to-interpretation representation. DL techniques offer the advantage of extracting the features from unprocessed data, but it might take much work to interpret the relative significance of the incoming data. Prior research

[27] established the concept of attention to solve this issue. This research introduced an attention mechanism for neural network machine translation operations into our classification system. This effort helped the construction of an interpretable representation of the model's input data components. As shown in Tables 6 and 7, the findings revealed that the attention mechanism improved detection capability in all conditions. Surprisingly, the effectiveness of our SenPyramidNet model was much improved for indoor and outdoor motion identification.

Table 6: Improved performance of SenPyramidNet using attention mechanism for indoor activity recognition

Model	Performance (Mean (±Std.))		
SenPyramidNet	Accuracy	Loss	F1-score
Without attention	95.54% (±0.43%)	0.21 (±0.035)	95.22% (±0.51%)
With attention	96.37% (±0.35%)	0.17 (±0.01)	96.08% (±0.39%)

Table 7: Improved performance of SenPyramidNet using attention mechanism for outdoor activity recognition

| Model	Performance (Mean ± Std.)		
	Accuracy	Loss	F1-score
Without attention	97.06% (±0.28%)	0.18 (±0.03)	96.77% (±0.27%)
With attention	97.25% (±0.43%)	0.19 (±0.03)	97.95% (±0.47%)

5.3 Model Parameters

The complexity of DL models employed in studies has been investigated. Table 8 demonstrates that SenPyramidNet has considerably greater parameters than ResNet16 but significantly fewer than other DL models. As demonstrated in Fig. 8, our proposed model is more accurate than theirs in indoor and outdoor action identification.

Table 8: Model parameters

	Models										
	Inception-ResNet	Inception	Xception	VGG11	VGG13	VGG16	VGG19	ResNet16	ResNet18	ResNet34	Sen Pyramid Net
Parameter (M)	44.43	14.29	20.71	28.29	28.35	30.12	31.89	0.46	3.85	7.24	0.79

(a)

(b)
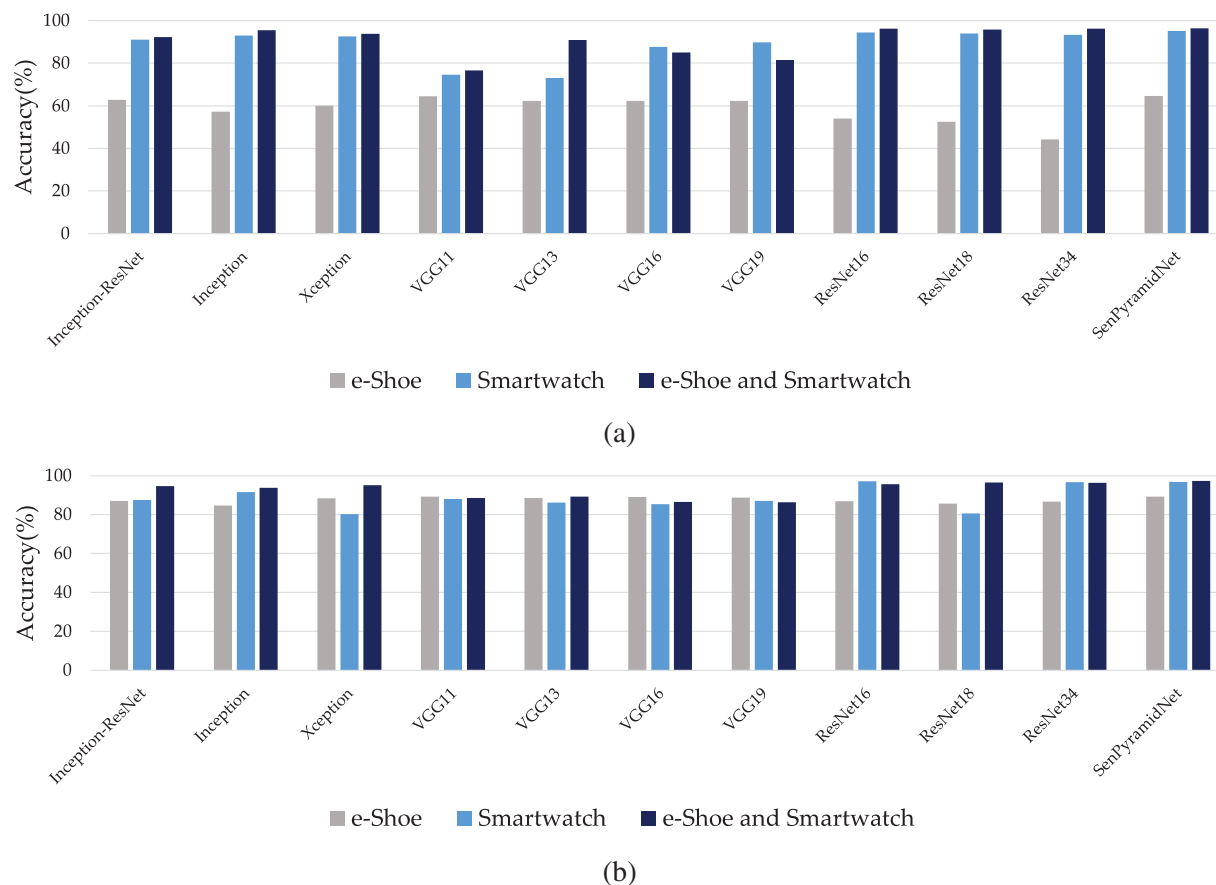
Figure 8: Comparison results with state-of-the-art DL models

6 Conclusion and Future Works

This research classified indoor and outdoor behaviors using DL algorithms. Various advanced DL approaches in HAR were evaluated by using the 19 Nonsens dataset. SenPyramidNet is a deep pyramidal residual network developed to efficiently distinguish indoor and outdoor actions based on sensor data from wearable devices. This model is a novel DL approach that combines the advantages of interaction modules with an attention mechanism to enhance the identification accuracy of the HAR. The SenPyramidNet model surpassed other state-of-the-art approaches concerning overall accuracy and F1-score. The indoor and outdoor activity identification accuracy rates of 96.37% and 97.25%, respectively, indicate the superiority of our model over existing sophisticated DL approaches. Comprehensive experiments have been carried out to investigate the performance of the proposed approach and demonstrate its robustness. The analyzed data indicate that the proposed SenPyramidNet recognizes indoor and outdoor activities effectively.

In the future, we will attempt to overcome one of the initial study's limitations: the requirement to collect sensor data with a predetermined size by using flexible data segmentation. In addition, we aim to build a pedagogical learning technique to enhance awareness of indoor and outdoor activities.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Mekruksavanich and A. Jitpattanakul, "Sport-related activity recognition from wearable sensors using bidirectional GRU network," *Intelligent Automation & Soft Computing*, vol. 34, no. 3, pp. 1907–1925, 2022.

[2] M. Zhang, S. Chen, X. Zhao and Z. Yang, "Research on construction workers' activity recognition based on smartphone," *Sensors*, vol. 18, no. 8, pp. 1–18, 2018.

[3] S. Mekruksavanich and A. Jitpattanakul, "LSTM networks using smartphone data for sensor-based human activity recognition in smart homes," *Sensors*, vol. 21, no. 5, pp. 1–25, 2021.

[4] A. Hayat, F. Morgado-Dias, B. P. Bhuyan and R. Tomar, "Human activity recognition for elderly people using machine and deep learning approaches," *Information*, vol. 13, no. 6, pp. 1–13, 2022.

[5] 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, pp. 1–19, 2020.

[6] J. Wang, Y. Chen, S. Hao, X. Peng and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, no. 4, pp. 3–11, 2019.

[7] X. Liu, L. Liu, S. J. Simske and J. Liu, "Human daily activity recognition for healthcare using wearable and visual sensing data," in *Proc. of ICHI*, Chicago, IL, USA, pp. 24–31, 2016.

[8] X. Yang and Y. Tian, "Super normal vector for human activity recognition with depth cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 1028–1039, 2017.

[9] M. Sharif, M. A. Khan, T. Akram, M. J. Younus, T. Saba *et al.,* "A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based features selection," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, pp. 1–18, 2017.

[10] C. M. Patil, B. Jagadeesh and M. N. Meghana, "An approach of understanding human activity recognition and detection for video surveillance using HOG descriptor and SVM classifier," in *Proc. of CTCEEC*, Mysore, India, pp. 481–485, 2017.

[11] X. Ji, C. Wang and Z. Ju, "A new framework of human interaction recognition based on multiple stage probability fusion," *Applied Sciences*, vol. 7, no. 6, pp. 1–16, 2017.

[12] Z. Wang, J. Wang, J. Xiao, K. H. Lin and T. Huang, "Substructure and boundary modeling for continuous action recognition," in *Proc. of CVPR*, Providence, RI, USA, pp. 1330–1337, 2012.

[13] Ö. F. İnce, I. F. Ince, M. E. Yıldırım, J. S. Park, J. K. Song *et al.,* "Human activity recognition with analysis of angles between skeletal joints using a RGB-depth sensor," *ETRI Journal*, vol. 42, no. 1, pp. 78–89, 2020.

[14] N. Irvine, C. Nugent, S. Zhang, H. Wang and W. W. Y. Ng, "Neural network ensembles for sensor-based human activity recognition within smart environments," *Sensors*, vol. 20, no. 1, pp. 1–26, 2020.

[15] Z. Feng, L. Mo and M. Li, "A random forest-based ensemble method for activity recognition," in *Proc. of EMBC*, Milan, Italy, pp. 5074–5077, 2015.

[16] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single tri-axial accelerometer," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1780–1786, 2014.

[17] B. M. H. Abidine, L. Fergani, B. Fergani and M. Oussalah, "The joint use of sequence features combination and modified weighted svm for improving daily activity recognition," *Pattern Analysis & Applications*, vol. 21, no. 1, pp. 119–138, 2018.

[18] F. De Cillis, F. De Simio and R. Setola, "Long-term gait pattern assessment using a tri-axial accelerometer," *Journal of Medical Engineering & Technology*, vol. 41, no. 5, pp. 346–361, 2017.

[19] Y. Tian, X. Wang, W. Chen, Z. Liu and L. Li, "Adaptive multiple classifiers fusion for inertial sensor based human activity recognition," *Cluster Computing*, vol. 22, no. 5, pp. 8141–8154, 2018.

[20] L. Jung and Z. Cheng, "Recognition of daily routines and accidental event with multipoint wearable inertial sensing for seniors home care," in *Proc. of SMC*, Banff, AB, Canada, pp. 2324–2389, 2017.

[21] A. R. Javed, M. U. Sarwar, S. Khan, C. Iwendi, M. Mittal *et al.,* "Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition," *Sensors*, vol. 20, no. 8, pp. 1–18, 2020.

[22] C. Pham, S. Nguyen-Thai, H. Tran-Quang, S. Tran, H. Vu *et al.,* "SensCapsNet: Deep neural network for non-obtrusive sensing based human activity recognition," *IEEE Access*, vol. 8, pp. 86934–86946, 2020.

[23] D. Anguita, A. Ghio, L. Oneto, X. Parra and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. of ESANN*, Bruges, Belgium, pp. 437–442, 2013.

[24] D. Han, J. Kim and J. Kim, "Deep pyramidal residual networks," in *Proc. CVPR*, Honolulu, HI, USA, pp. 6307–6315, 2017.

[25] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners.* Berkeley, CA, USA: Apress Berkeley, 2019.

[26] D. P. Kingma and J. Ba, "ADAM: A method for stochastic optimization," in *Proc. of ICLR*, San Diego, CA, USA, pp. 1–15, 2014.

[27] T. Luong, H. Pham and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. of EMNLP*, Lisbon, Portugal, pp. 1412–1421, 2015.