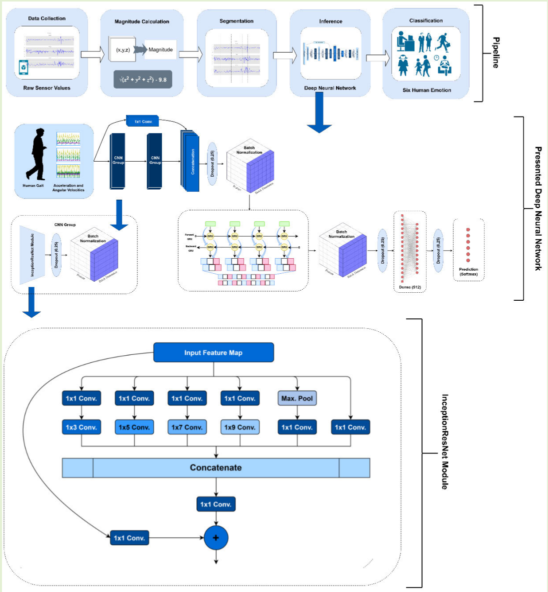
Smart-Wearable Sensors and CNN-BiGRU Model: A Powerful Combination for Human Activity Recognition

Hamza Ali Imran, Qaiser Riaz, *Senior Member, IEEE*, Mehdi Hussain, *Senior Member, IEEE*, Hasan Tahir, *Senior Member, IEEE*, and Razi Arshad, *Senior Member, IEEE*

***Abstract*—Human activity recognition (HAR) is a key component of ambient-assisted living and one of the most active areas of research in the Internet of Things (IoT). The use of wearable and embedded sensors in HAR overcomes the limitations of conventional approaches relying on machine vision and environmental sensors. We offer a novel, lightweight convolutional neural network–bidirectional gated recurrent unit (CNN-BiGRU) model that classifies human activities using the inertial sensor data collected with body-mounted smartwatches and smartphones. Unlike the traditional approaches, the presented model is trained on the magnitude of the 3-D acceleration ($\widehat{mag_a}$), which significantly minimizes the input space 1-D. The deep learner has been validated using two different publicly available datasets from the wireless sensor data mining (WISDM) lab and different evaluation parameters, such as recall/sensitivity, precision, accuracy, and F1-scores, are computed. A comparison with the existing studies reveals that our proposed learner surpasses the existing methodologies. Using magnitude of 3-D acceleration ($\widehat{mag_a^w}$, 1-D input signal), we have achieved 97.29% accuracy for all six activities of WISDM 2011 dataset and 98.81% accuracy for 3-D acceleration (a_x^w, a_y^w, and a_z^w). The precision, recall, and F1-score remained at 97% for the 1-D case and 99% for the 3-D case. When evaluated on the data of all 18 smartwatch-based activities in the WISDM 2019 dataset, we have achieved 97.5% accuracy with the magnitude of 3-D acceleration ($\widehat{mag_a^w}$) and 98.4% accuracy for 6-D acceleration and angular velocities (a_x^w, a_y^w, a_z^w, ω_x^w, ω_y^w, and ω_z^w). The precision, recall, and F1-score remained at 98% in both cases.**



***Index Terms*— Deep models for activity recognition, gait analysis, human activity recognition (HAR), inertial sensors, signal processing, wearable sensors.**

I. Introduction

THE Internet of Things (IoT) envisions a future in which common things are outfitted with microcontrollers, wireless communication transceivers, and protocol stacks. The ability to sense the surroundings and the human body's connection with the physical world are critical and serve as major features of the IoT world. Wearable inertial sensors, e.g., smartwatches, smartphones, and smart fitness bands, can be easily attached to the human body and can store variable frequencies of low-level human body kinematics in real time. The same concept is applied to many application areas such as human activity recognition (HAR) [1], fall detection [2], identifying and understanding human emotions [3], [4], human joints pattern identification [5], and person re-identification [6]. The use of wearable inertial sensors for HAR, in particular, is a strong and efficient solution for ambient-assisted living, and as artificial intelligence (AI) advances, it is getting increasingly accurate and independent.

HAR is an active area in the sphere of the IoT [1]. HAR applications include, but not limited to, health monitoring

Manuscript received 1 June 2023; revised 21 November 2023; accepted 28 November 2023. Date of publication 7 December 2023; date of current version 12 January 2024. The associate editor coordinating the review of this article and approving it for publication was Dr. Qammer H. Abbasi. *(Corresponding author: Qaiser Riaz.)*

The authors are with the Department of Computing, School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan (e-mail: himran.mscs18seecs@seecs.edu.pk; qaiser.riaz@seecs.edu.pk; mehdi.hussain@seecs.edu.pk; hasan.tahir@seecs.edu.pk; razi.arshad@seecs.edu.pk).

Digital Object Identifier 10.1109/JSEN.2023.3338264

and recreation, fitness tracking, ambient supported living, surveillance, patient monitoring, and many others [7], [8]. The HAR problem is challenging in nature and researchers have approached it in different ways, e.g., using vision sensors, environmental sensors, and wearable sensors. Occlusion, illumination variations, operational expenses, privacy concerns, and other factors limit vision-based systems [9], [10]. Environmental sensor-based techniques typically overcome the limits associated with visual sensors, but they cannot be used in settings requiring constant monitoring, such as the care of senior individuals [11]. For continuous monitoring, wearable sensors have minimal limitations and they also provide better privacy, which is usually compromised in vision-based and environmental sensor-based approaches. For noninvasive, on-body sensor placements to characterize human activities, inertial sensors have been employed in several studies for data collection and analysis [12], [13], [14], [15], [16], [17], [18]. In these studies, it has been observed that the low-level kinematic data can be used in training learners to predict different types of locomotive activities—walking, jogging, and running—as well as nonlocomotive activities—standing, sitting, eating, and so on.

In the context of HAR, there exist several datasets such as wireless sensor data mining (WISDM) 2011 [12], WISDM 2019 [13], SMARTPHONE/UCI-HAR [14], MHEALTH [16], and OPPORTUNITY [15]. Out of these datasets, the WISDM 2011 dataset [12] is an extensive dataset that includes the inertial data of six different human activities collected using body-mounted smartphones under the controlled environment. The WISDM 2019 dataset [13] is an extension of previously collected data and it contains the inertial data of 18 different activities collected using wrist-attached smartwatches in the open environment (outside of the lab). Since the data are collected under controlled/uncontrolled environments, it makes the WISDM datasets quite challenging to analyze and predict.

In this work, we proposed a novel deep learning network, which has two multitype convolutional kernels with residual connections and bidirectional gated recurrent unit (BiGRU) units. It classifies diverse daily life activities using the raw inertial sensor data (WISDM 2011 and WISDM 2019 publicly available datasets for training and validation). Most of the existing approaches solve the HAR problem by separately feeding 6-D acceleration and angular velocities into the deep networks in order to predict daily activities. Because of 6-D signals, the deep learner requires extensive computational resources for training resulting in higher computation time. However, the proposed approach significantly differs from the existing approaches as it relies on the 1-D magnitude of triaxial accelerations ($\widehat{\text{mag}_a}$) only. The magnitude is computed from the raw input signal during the preprocessing phase. Due to a significant decrease in the input size to 1-D ($\widehat{\text{mag}_a}$), the requirement of computation resources is much lower, and hence, the processing time is decreased remarkably. Moreover, the proposed approach not only minimizes the overall complexity of the model but also surpasses existing methodologies. The significant contributions are given as follows.

1) We propose a convolutional neural network–BiGRU (CNN-BiGRU) model, in which we introduce a novel concept of the raw link, which is actually a 1 × 1 convolution link from the raw input to the BiGRU model (refer to Section IV).
2) To reduce the processing requirements, we utilize the magnitude of 3-D acceleration ($\widehat{\text{mag}_a}$) to minimize the input space to 1-D (refer to Section IV).
3) We train and test the proposed model using different sets of input signals (1-D–6-D) for comparison purposes, and we report the best results obtained (refer to Section V).
4) We empirically evaluate the impact of segmentation window size and step size on sequential data of signals (refer to Section V-C).
5) The performance of the proposed learner is compared with the existing ones on WISDM datasets and it is observed that it surpasses the state-of-the-art methodologies (Section VI-A).

The remaining article is divided so that Section II presents the related literature and Section III shares the details of datasets used for evaluation. Section IV covers details of preprocessing, presented model, training process, and complexity of the model. Section V covers the details of the evaluation of the proposed model and Section VI concludes this article.

II. Literature Review

Zhang et al. [19] presented a deep learning model similar to Inception-Net, where three different sized kernels feature extractors are used, and afterward, attention mechanisms are applied. Using the WISDM 2011 dataset, they reported an F1-score of 94.5%. Peppas et al. [20] proposed two CNN models for HAR. They used windows of different sizes (50–200) to train the models. They tested the models on the WISDM 2011 dataset and the Actitracker dataset [21]. Xia et al. [22] suggested a design where long short-term memory (LSTM) layers are followed by CNN layers followed by a global average pooling. The model was evaluated using three datasets SMARTPHONE/UCI-HAR, WISDM 2011, and OPPORTUNITY. The classification accuracies were 95.78%, 95.85%, and 92.63%. Imran and Latif [9] presented a CNN, which has dense connections between inception-like modules having three different sized kernels with global average pooling. It was evaluated on SMARTPHONE/UCI-HAR and WISDM lab 2011 datasets. The average F1-score for the WISDM 2011 dataset was 95%. Mehmood et al. [10] proposed a DenseNet-inspired architecture for HAR. They evaluated the model on three datasets SMARTPHONE/UCI-HAR, WISDM lab 2011, and WISDM Actitracker datasets and have achieved 91.6%, 94.65%, and 77.18%, respectively.

Dua et al. [23] presented various models for HAR, where a deep convolutional model with a random forest classifier worked the best. The presented models are evaluated on WISDM 2011 and SMARTPHONE/UCI-HAR datasets and achieved the accuracies of 97.77% and 98.2%, respectively. A feature set enhancing mechanism is introduced in the study [24] where the input feature maps are passed through CNN followed by a reverse operation for deconvolutions and so on. The output is then combined with the input feature map, thus increasing the training feature map size. They tested their approach on five openly available HAR datasets; for the WISDM 2011 dataset, they reported 99%, 99.03%, and 99.03% classification accuracies for CNN with original, ten times combined features, and 100 times combined features, respectively. For CNN-LSTM, they got 99.41%, 99.51%, and 99.47% accuracy for the same cases. Nafea et al. [25] presented a two-stream deep model using CNN and BiLSTM. The features extracted by the two networks are fused and used for classification. The presented approach is evaluated on the SMARTPHONE/UCI-HAR and WISDM 2011 datasets and the classification accuracies of 97.05% and 98.53% are reported. Imran [26] presented an antithesis kind of model having bidirectional GRU units followed by CNNs evaluated on the WISDM 2011 dataset and they reported that the classification accuracy is 97.2%.

Weiss et al. [13] have presented the dataset, which is referred to in this study as WISDM 2019 dataset. The team has worked on activity recognition as well as inertial sensor-based biometrics identification. The accuracy achieved for HAR was 94.4%. Qin et al. [27] suggested a multichannel learner based on CNN + BiLSTM and evaluated it on two datasets: WISDM 2019 and MHEALTH. On the WISDM dataset, they reported a weighted averaged F1-score of 99%, while on the MHEALTH dataset, they reported a weighted averaged F1-score of 100%. They combined all of the axes values from the mobile's and the smartwatch's accelerometer and gyroscope sensors for the WISDM dataset, making their solution much more complex.

III. Datasets

We have selected two publicly available datasets from the WISDM lab for training, validation, and prediction of human activities. The data are collected in both controlled and unconstrained environments using different sensor modalities such as smartphones and smartwatches.

A. WISDM 2011

The dataset was collected in 2011 using on-body smartphones. A total of 29 volunteer subjects were hired and they were instructed to perform different activities while carrying the smartphone in the front pocket of their pants. The data of the built-in accelerometers were recorded. The dataset consists of a total of "1 098 207" samples covering six daily activities, i.e., "jogging," "walking," "upstairs," "downstairs," "upstairs," "standing," and "sitting." Fig. 1 reveals the class distribution for various activities.

B. WISDM 2019

The dataset contains 18 different activities of daily living. Sensory data from smartphones and smartwatches were obtained from 51 volunteers (age: 18–25 years). The dataset consists of low-level signals recorded from the 3-D accelerometer and 3-D gyroscope. For this study, we have used the inertial data collected with the smartwatches to train and validate the proposed model. The choice of smartwatch's data also helps us in evaluating the effect of change in the orientation of the device, which is quite prominent when the device is attached to the subject's dominant wrist. Fig. 2 shows the class distribution of different activities for this dataset.

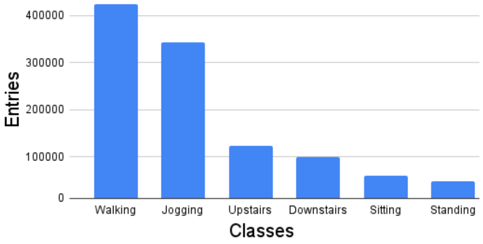


Fig. 1. Class distribution of WISDM 2011 dataset.

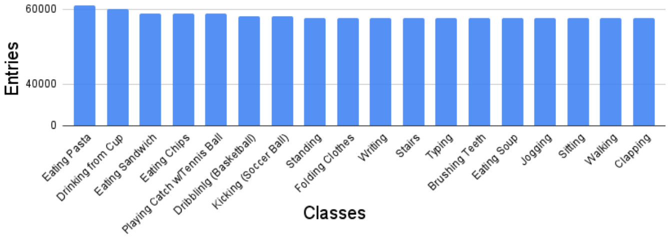


Fig. 2. Class distribution of WISDM 2019 dataset.

IV. Methodology

In this section, the proposed methodology is described. The section starts with the preprocessing (Section IV-A) where we compute the magnitude of 3-D acceleration. The architecture of the proposed methodology is discussed in (Section IV-B). Before concluding the section, we discuss the training process in detail (Section IV-C). The end-to-end flow of the work at hand is shown in Fig. 3.

A. Preprocessing

Our work primarily aims to increase computational efficiency, making it better suited for real-time systems and IoT applications. One way to do this is to decrease the input size for [28]. Thus, in order to translate the input from 3-D to 1-D, we propose computing the magnitude of 3-D acceleration. The following formula is used to determine the size of the 3-D acceleration

$$\widehat{\mathrm{mag}_a} = \sqrt{\left(a_x^2 + a_y^2 + a_z^2\right)} - 9.8. \tag{1}$$

We calculate $\widehat{\mathrm{mag}_a}$ during the preprocessing phase and use it as a separate input signal in the deep neural network for training and validation. The input dimensionality is effectively decreased from 6-D to 1-D as a result. Utilizing magnitude is primarily done to lessen the effect of sensor orientation. Particularly when the sensor is fixed on a joint (such as a smartwatch on the subject's dominant wrist), sensor orientation has a significant impact.

B. Architecture

We propose HARDenseRNN a model to recognize human activity that combines CNN and recurrent neural network (RNN) components. Following a 128-unit bidirectional GRU, our model consists of two CNN modules with multitype kernels and residual connections. The CNN feature map and the raw data are combined using a 1 × 1 convolutional layer connection. The reasoning for adding raw data is that if

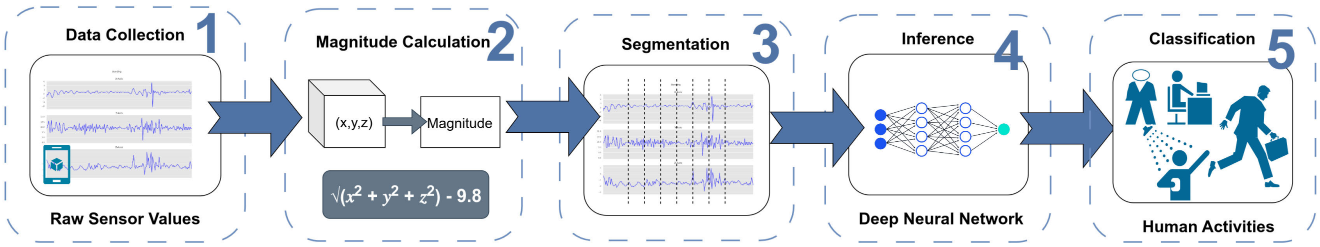


Fig. 3. Pipeline of the proposed approach consists of five steps. The magnitude of the sensor's raw data is computed, segmented, and fed into the proposed HARDenseRNN model for inference.

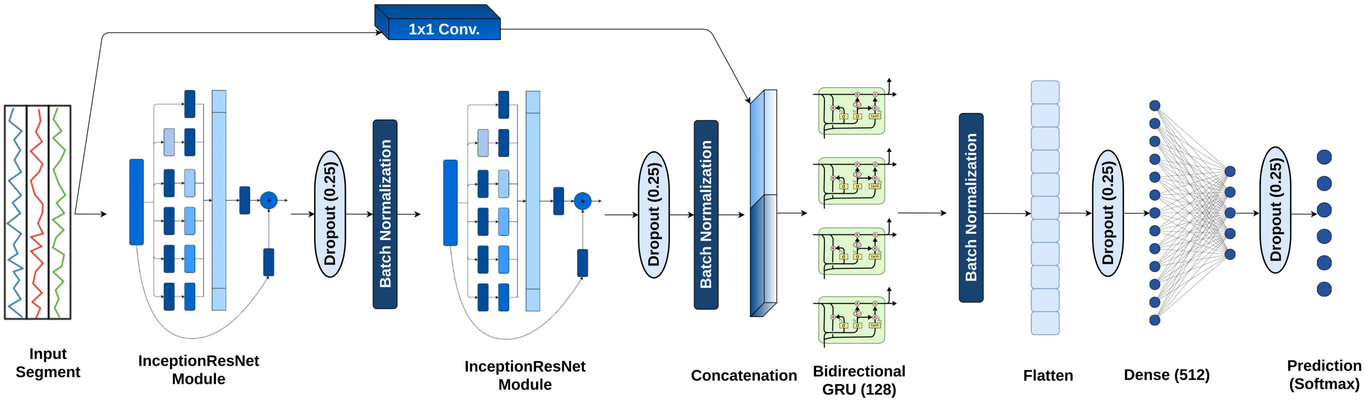


Fig. 4. Overall model is depicted in this diagram schematically. Two multikernel CNN modules receive the input vector. The computed feature map is combined with the low-level signals and sent to a 128-unit bidirectional GRU for batch normalization. The generated vector is flattened before being sent to the fully connected layer (512 neurons). The dropout layer assures regularization.

CNNs do not significantly improve the classification accuracy, the low-level sensor data are still available for the RNNs. To maintain the spatial dimensions, we use 1 × 1 convolution techniques. We have used 64 (1 × 1) kernels for convolution on the raw sensor data. Our experiments demonstrate a significant improvement in the performance of the model by the introduction of the raw data.

Utilizing multiple types of kernels enables the detection of features of various sizes. The model can learn identity mappings due to the incorporation of residual connections, allowing it to skip over some modules that do not have a big impact on classification. RNNs typically outperform other models when processing sequential data, and hence, the bidirectional GRU receives the feature map from the CNN modules. The overall proposed architecture is shown in Fig. 4, whereas CNN modules incorporated into the model are shown in Fig. 5.

The kernels in the proposed CNN module have five different sizes: 1 × 1, 1 × 3, 1 × 5, 1 × 7, and 1 × 9. The input is convoluted with ten 1 × 1 kernels before applying ten 1 × 3, 1 × 5, 1 × 7, and 1 × 9 kernels. The overall complexity of the model is decreased by this approach and a similar approach is presented in [25]. We investigated adding kernels of various sizes sequentially, and the results showed performance gains. Based on empirical findings, the kernels and spatial dimensions were selected (see Section V-C). We then have a concatenation of the feature maps produced by the kernels and max pooling, which significantly enhances the performance. Afterward, the spatial dimensions of the resulting feature map are reduced by running it through 64 1 × 1 convolutional kernels. In parallel, the 64, 1 × 1 kernel convolution of the input feature map is also carried out and the resulting feature maps from both branches are combined. Based on empirical analysis, we have used a pair of these modules in the proposed architecture. After each module, batch normalization and a dropout = 0.25 are applied. A fully connected layer with 512 neurons receives the flattened and reshaped feature vector produced by the BiGRU units. For the purposes of regularization, a dropout of 0.25 is used.

C. Training Process

All of the experiments were conducted at the Google Colaboratory, and runtime GPU was chosen. Tensorflow 2.4.0 and Keras 2.4.3 were used as the backends when creating the model. The validation process used a tenfold cross-validation strategy along with the following: batch size = 64, number of epochs = 120, and optimizer = Adam.

V. Results

The results are presented using confusion matrices, accuracy, precision, recall, and F1-score. Since the WISDM 2019 dataset is fairly balanced, accuracy is a solid indication of the performance of the model, whereas the WISDM 2011 dataset is highly unbalanced; hence, accuracy is not a desirable performance measure, so the other evaluation parameters are also computed. This gives us a realistic and true picture of the model's performance and behavior

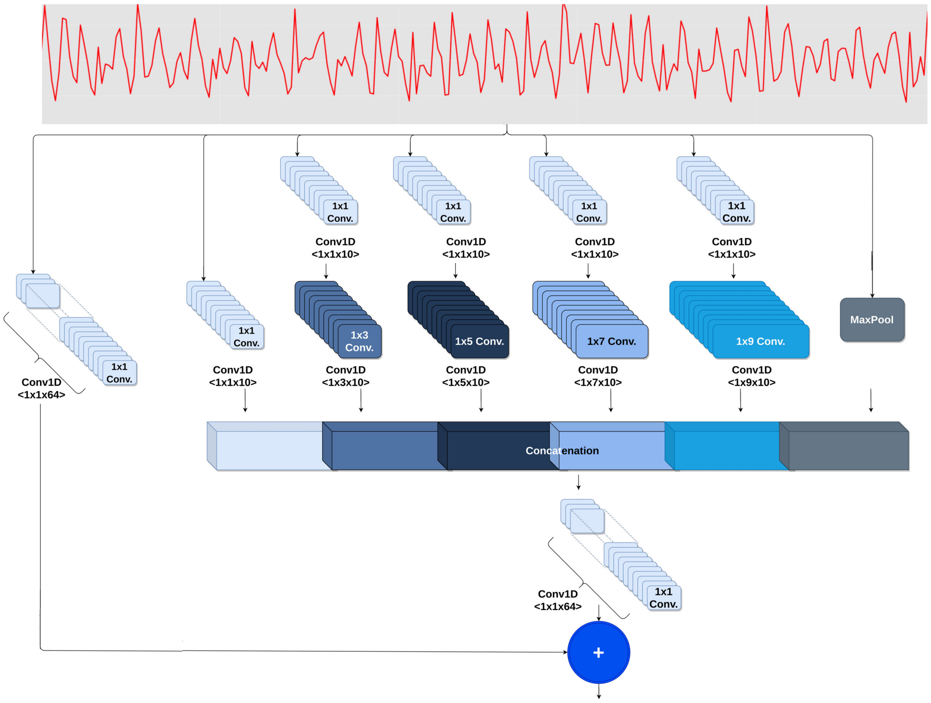


Fig. 5. CNN module proposed in this work involves max pooling and convolving the input feature map with kernels of five different sizes. Before using kernels of various sizes, a 1 × 1 convolution is first applied to reduce overall complexity. In order to add the resultant feature maps to the initial feature map, concatenation is used. In addition, 1 × 1 convolutions are used to guarantee uniform spatial dimensions.

A. HAR Using Smartphone Data

We compute the magnitude of raw 3-D acceleration and remove the gravitational constant from it. This 1-D signal is then fed into the proposed deep model. For the sake of comparison, we also compute the results of 3-D accelerations. The results are reported in the following.

1) Magnitude of 3-D Acceleration: Table I shows the comparison of precision, F1-score, and recall achieved with the proposed model using the magnitude of 3-D acceleration as input. All of the three parameters remain at 97%. The lowest scores for precision, recall, and F1-score are observable for the *sitting* and *standing* activities. This is because both of these activities are nonlocomotive and only minimal inertial impact is present in such nonlocomotive activities. The confusion between the *sitting* and *standing* activities is observable in the confusion matrix [see Fig. 6(a)] where 15% of the *sitting* activities are confused with the *standing* activities and 16% of the *standing* activities are confused with the *sitting* activities. The average accuracy remains above 97%.

2) 3-D Acceleration: Table I shows the comparison of precision, recall, and F1-score achieved with the proposed model using the raw 3-D acceleration as input. All of the three parameters remain at 99%. The performance has been increased in this case, which was expected, but the complexity of the model is also increased. The input size is increased from 1-D to 3-D. The lowest scores in this case are not of *sitting* and *standing* but are for *upstairs* and *downstairs*. The performance for *upstairs* and *downstairs* is not improved but remained the same, but a considerable improvement can be seen for *sitting* and *standing*, which results in the overall improvement in accuracy by 1.52%. The achieved average accuracy is 98.81%. It is evident from the confusion matrix [see Fig. 6(b)] that 2.6% of the *downstairs* activities are confused with the *upstairs* activities and 2.6% of the *upstairs* activities are confused with the *downstairs* activities.

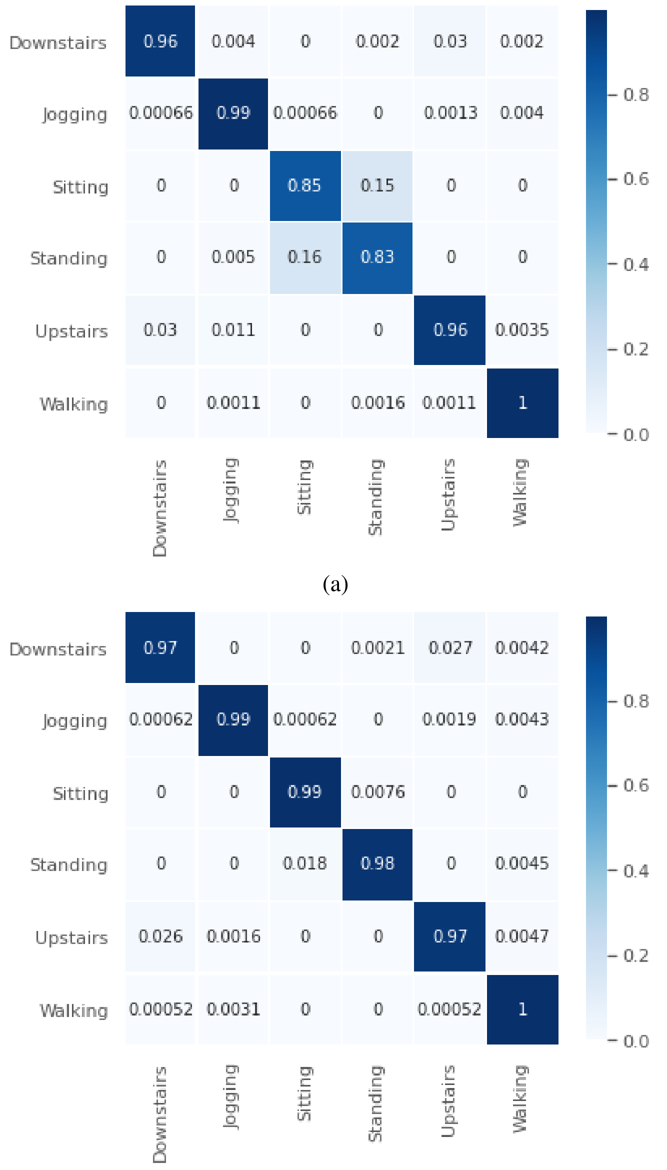
TABLE I
PRECISION, F1-SCORE, RECALL/SENSITIVITY, AND SPECIFICITY COMPUTATIONS AGAINST VARIOUS INPUT SIGNALS (1-D-TO-3-D) USING THE WISDM 2011 DATASET ARE COMPARED

Input	$\widehat{mag_a^p}$	a_x^p, a_y^p, a_z^p	$\widehat{mag_a^p}$	a_x^p, a_y^p, a_z^p
Activities	**Precision (%)**		**Recall/Sensitivity (%)**	
Downstairs	96	96	96	97
Jogging	99	100	99	99
Sitting	86	98	85	99
Standing	80	99	83	98
Upstairs	97	97	96	97
Walking	100	99	100	100
Activities	**F1-Score (%)**		**Specificity (%)**	
Downstairs	96	96	98.56	97.03
Jogging	99	99	98.89	99.26
Sitting	85	99	99.27	99.26
Standing	82	98	98.88	98.32
Upstairs	96	97	98.71	97.22
Walking	100	99	99.17	96.94

3) Performance Comparison of Different Cases: A tradeoff has been noted between the performance of the model with complexity. We have compared both the above-discussed cases through a comparison Table II.

B. HAR Using Smartwatch Data

We have experimented with five different cases, which include using the magnitude of 3-D acceleration (1-D input signal), the magnitude of 3-D angular velocities (1-D input signal), 3-D acceleration, 3-D angular velocities, and 6-D



(a)

(b)

Fig. 6. Confusion matrix of classification based on the WISDM 2011 dataset using smartphone's accelerometer. (a) Magnitude of 3-D acceleration. (b) 3-D acceleration.

TABLE II

Performance Comparison Between Magnitude of 3-D Acceleration ($\widehat{mag}_a^p$) Versus 3-D Acceleration (a_x^p, a_y^p, and a_z^p) Used as Input Signal to the Proposed Model. The Model Is Trained and Validated With the WISDM 2011 Dataset Collected Using a Smartphoneś Triaxial Accelerometer

Input	Accuracy (%)	F1-Score (%)	Precision (%)	Recall/Sensitivity (%)
$\widehat{mag}_a^p$	97.29	97.00	97.00	97.00
a_x^p, a_y^p, a_z^p	98.81	99.00	99.00	99.00

acceleration and angular velocities for only smartwatch data. In the following, we discuss our findings of the model when it is trained and validated with: 1) the magnitude of 3-D acceleration; 2) 3-D acceleration; and 3) 6-D acceleration and angular velocities.

1) Magnitude of 3-D Acceleration: Table III shows the precision, recall, and F1-score achieved with the proposed model using the magnitude of 3-D acceleration, a 1-D signal. The evaluation parameters remain at 98%. The most challenging activities for the model are *sitting*, *standing*, *drinking from cup*, and *eating sandwich*. Considering the confusion matrix shown in Fig. 7, we can see that 1.1% *sitting* activities are confused with *standing* and 2.8% activities of *standing* are confused with *sitting*. These confusions are due to the similar nature of input signal generated with *sitting* and *standing* activities (i.e., both being nonlocomotive activities). Similarly, confusion between *drinking from cup* and *eating sandwich* are also observable where the average precision remains at 95%.

2) 3-D Acceleration: We have also experimented with the 3-D acceleration by feeding them as input to the proposed model. This case produces slightly better results, however, with higher computation cost (because of 3-D input size instead of the 1-D input size). With tenfold cross validation, the average accuracy of 98.2% is achieved. The precision, F1-score, and recall values remain at 98% (Table III). Fig. 8 presents the confusion matrix, and here, it is observable that most of the confusion exists between eating activities, i.e., *eating pasta* and *eating chips*.

3) 6-D Acceleration and Angular Velocities: The last case we experimented on is to feed 6-D acceleration and angular velocities into the proposed model and analyze the impact. This case is important as with the 6-D input signal, the computational complexity is multifold higher than the 1-D or 3-D input signal. Fig. 9 presents the confusion matrix and with the tenfold cross validation, and the average accuracy remains at 98.4%. The precision, F1-score, and recall also remain at 98%, as shown in Table III.

4) Performance Comparison of Different Cases: We have tested with five different types of input signals ranging from the 1-D signal (magnitude) to 3-D signals (3-D acceleration and 3-D angular velocities) to 6-D signals (6-D acceleration and angular velocities). Table IV presents a comparison of average accuracies, F1-score, recall, and precision between all of the five cases. From Table IV, it is observable that the magnitude of angular velocities ($\widehat{\text{mag}}_\omega^w$) performs worst where the accuracy remains at 63.9%. Similarly, the values of precision, F1-score, and recall remain lower, i.e., between 63% and 67%. The average classification accuracies for the remaining cases remain higher, i.e., between 96.8% and 98.4%. The trend also remains the same for the precision, F1-score, and recall, i.e., between 97% and 98%. It is important to note that the computational cost of the 1-D signal ($\widehat{\text{mag}}_a^w$) remains much lower than that of the 3-D signals (a_x^w, a_y^w, a_z^w) or 6-D signals ($a_x^w, a_y^w, a_z^w, \omega_x^w, \omega_y^w$, and ω_z^w) without any notable degradation in the average classification accuracy. Thus, it can be concluded that the 1-D signal, i.e., $\widehat{\text{mag}}_a^w$, can yield similar classification accuracies with much lower computational cost, which is a key limitation in most of the existing approaches.

C. Effect of Segmentation Window Size

The segmentation window size of the signal is an important factor in the classification of daily activities as variations in window sizes affect the classification accuracy.

TABLE III

PRECISION, F1-SCORE, RECALL/SENSITIVITY, AND SPECIFICITY COMPUTATIONS AGAINST VARIOUS INPUT SIGNAL (1-D-TO-6-D) USING THE WISDM 2019 DATASET ARE COMPARED. 3-D Acc = a_x^w, a_y^w, a_z^w, 3-D Gyro = $\omega_x^w, \omega_y^w, \omega_z^w$, AND 6-D Acc + Gyro = $a_x^w, a_y^w, a_z^w, \omega_x^w, \omega_y^w, \omega_z^w$

Input	$\widehat{mag_\omega^w}$	$\widehat{mag_\omega^w}$	**3D Acc**	**3D Gyro**	**6D Acc+Gyro**	$\widehat{mag_\omega^w}$	$\widehat{mag_\omega^w}$	**3D Acc**	**3D Gyro**	**6D Acc+Gyro**
Activities	**Precision (%)**					**Recall/Sensitivity (%)**				
Walking	98	98	99	99	99	99	97	97	99	96
Jogging	99	99	98	99	99	99	98	99	99	100
Stairs	97	80	99	98	98	98	94	99	98	100
Sitting	95	57	98	90	99	96	5.2	98	97	98
Standing	97	41	99	96	100	95	17	96	94	99
Typing	98	93	97	99	100	98	4.4	100	89	99
Brushing Teeth	99	45	99	99	97	99	97	99	100	99
Eating Soup	95	78	96	98	98	99	70	100	96	97
Eating Chips	99	30	98	95	96	98	72	97	93	99
Eating Pasta	99	32	98	97	100	98	57	94	96	97
Drinking from Cup	97	55	98	96	99	95	75	99	97	95
Eating Sandwich	97	62	98	92	97	95	30	98	94	98
Kicking (Soccer Ball)	98	82	99	99	98	99	99	100	97	98
Playing Catch w/Tennis Ball	99	99	99	97	98	98	99	99	100	98
Dribbling (Basketball)	99	96	99	99	97	100	99	99	100	99
Writing	99	98	99	95	100	99	10	100	99	100
Clapping	98	97	99	99	100	99	99	99	99	99
Folding Clothes	97	98	98	99	96	97	78	99	99	99
Activities	**F1-Score (%)**					**Specificity (%)**				
Walking	98	97	98	99	97	100	97.89	97.0	99	95.75
Jogging	99	98	99	99	99	99.66	97.41	99.0	99	100
Stairs	98	86	99	98	99	99.49	94.43	99.0	98	100
Sitting	96	9.5	98	93	99	99.45	73.16	98.0	97	95.04
Standing	96	24	97	95	99	99.44	78.72	96.0	94	99.04
Typing	98	8.3	98	94	99	99.45	61.16	100	89	100
Brushing Teeth	99	62	99	99	98	99.34	98.55	99	100	100
Eating Soup	97	74	98	97	97	99.34	84.16	100	96	94.29
Eating Chips	99	43	97	94	97	99.49	90.07	97	93	100
Eating Pasta	98	41	96	97	99	99.49	78.87	94	96	96.15
Drinking from Cup	96	64	99	97	97	99.57	96.05	99	97	91.89
Eating Sandwich	96	41	98	93	97	99.57	76.92	98	94	96.55
Kicking (Soccer Ball)	98	89	99	98	98	99.47	99.37	97	97	95.04
Playing Catch w/Tennis Ball	99	99	99	98	98	99.49	99.02	100	100	93.31
Dribbling (Basketball)	99	89	99	99	98	100	98.74	99	99	95.88
Writing	99	19	99	97	100	99.47	77.17	100	99	100
Clapping	98	98	99	99	99	98.75	98.87	99	99	100
Folding Clothes	97	87	98	99	98	99.43	85.80	99	99	99.01

TABLE IV

PERFORMANCE COMPARISON BETWEEN DIFFERENT TYPES AND DIMENSIONS OF INPUT SIGNAL USED IN THE PROPOSED MODEL. THE MODEL IS TRAINED AND VALIDATED WITH WISDM 2019 DATASET COLLECTED USING SMARTWATCHŚ TRIAXIAL ACCELEROMETER AND TRIAXIAL GYROSCOPE

Input	**Accuracy (%)**	**F1-Score (%)**	**Precision (%)**	**Recall/Sensitivity (%)**
$\widehat{mag_a^w}$	**97.5**	**98**	**98**	**98**
$\widehat{mag_\omega^w}$	63.9	63	75	67
(a_x^w, a_y^w, a_z^w)	**98.2**	**98**	**98**	**98**
$(\omega_x^w, \omega_y^w, \omega_z^w)$	96.8	97	97	97
$(a_x^w, a_y^w, a_z^w,$ $\omega_x^w, \omega_y^w, \omega_z^w)$	**98.4**	**98**	**98**	**98**

TABLE V

TEST ACCURACY COMPARISON OF DIFFERENT SEGMENTATION WINDOW SIZES AND STEP SIZES

Window Size	**Step Size**	**Overlapping Samples (%)**	**Test Accuracy (%)**
64	16	75.00	83.72
64	32	50.00	72.51
128	32	75.00	91.12
128	64	50.00	81.64
256	16	93.75	99.36
256	32	87.50	97.71
256	64	75.00	94.60
256	128	50.00	86.57

Therefore, we have experimented with different segmentation strategies.

The size of the segmentation window was initially set to 256 along with the stepping size of 64. Once converged to the presented network, we explored the effect of segmentation window size. For the network hunt, we divided the dataset into three sets of 70% training, 15% testing, and 15% validation.

The stepping size is the number of samples the window is slid and before the second segment is created. Table V presents the effect of different segment window sizes along with step size and the test data accuracies are reported. For this experiment, the dataset was divided into two chunks, i.e., 90% for training and 10% for testing. All these experiments were conducted with the magnitude of the accelerometer of the smartwatch, i.e., $\widehat{\mathrm{mag}_a^w}$ [see (1)]. From the experimentation results, we choose the following best parameters: sampling window size = 256 and step size = 32. It is observable from

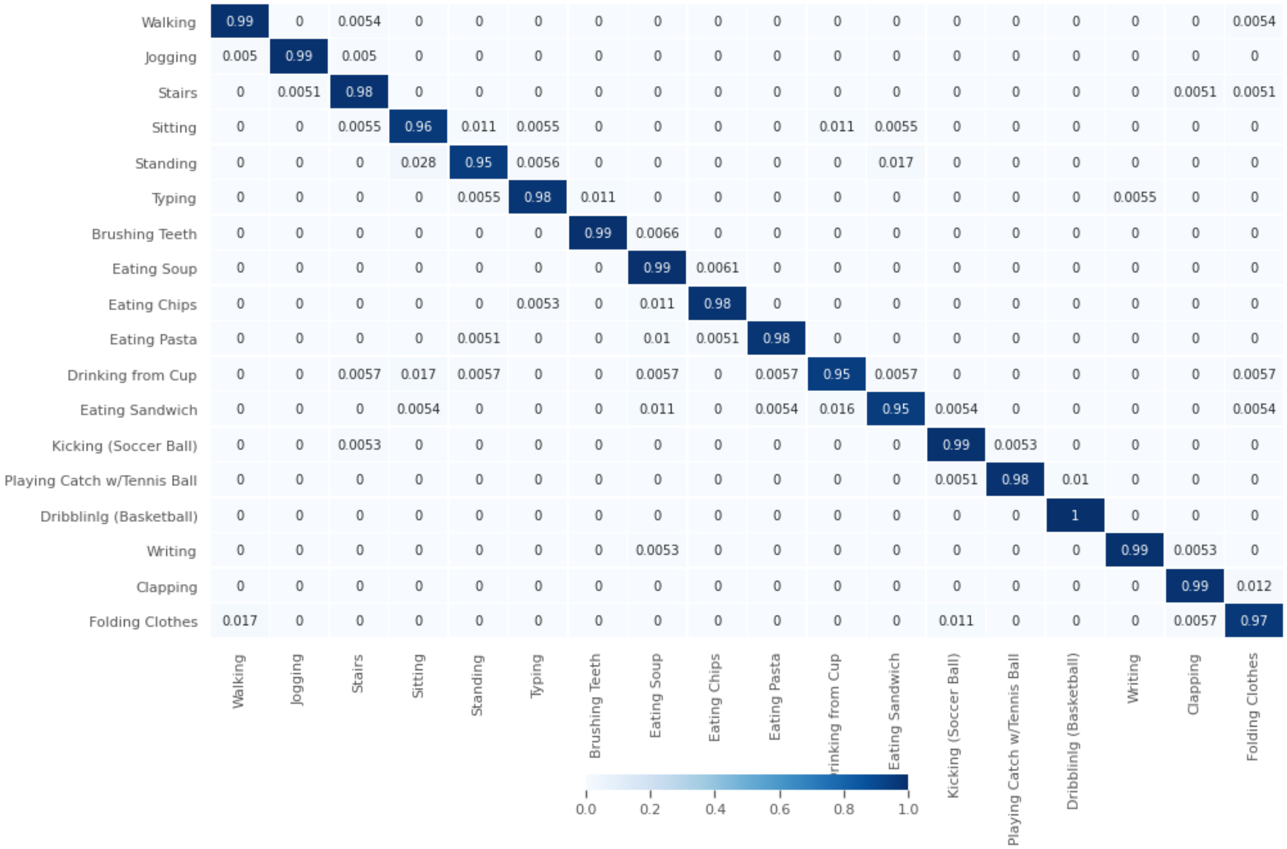


Fig. 7. Confusion matrix computed with the magnitude of 3-D acceleration ($\widehat{\text{mag}_a^w}$) using the WISDM 2019 dataset (smartwatch data). The average classification accuracies remain at 97.5%.

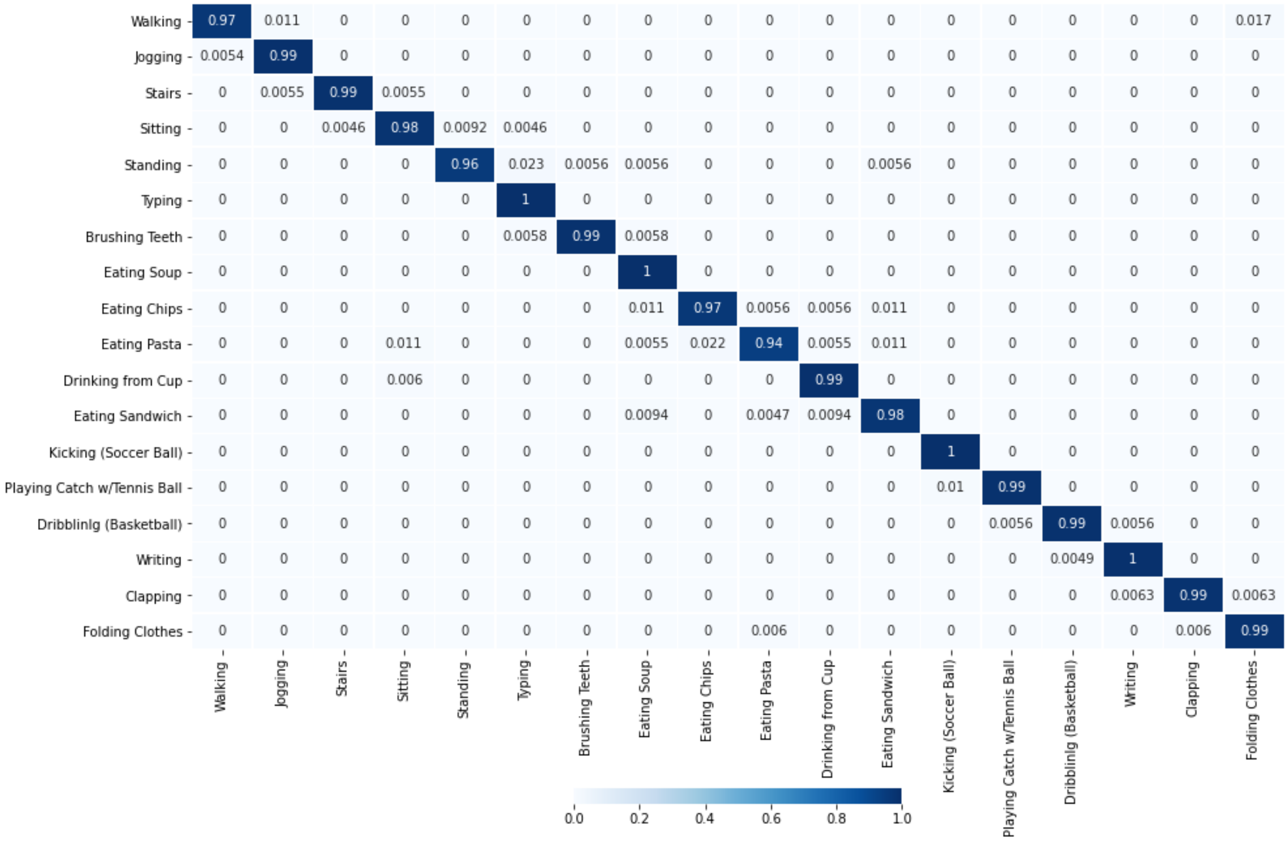
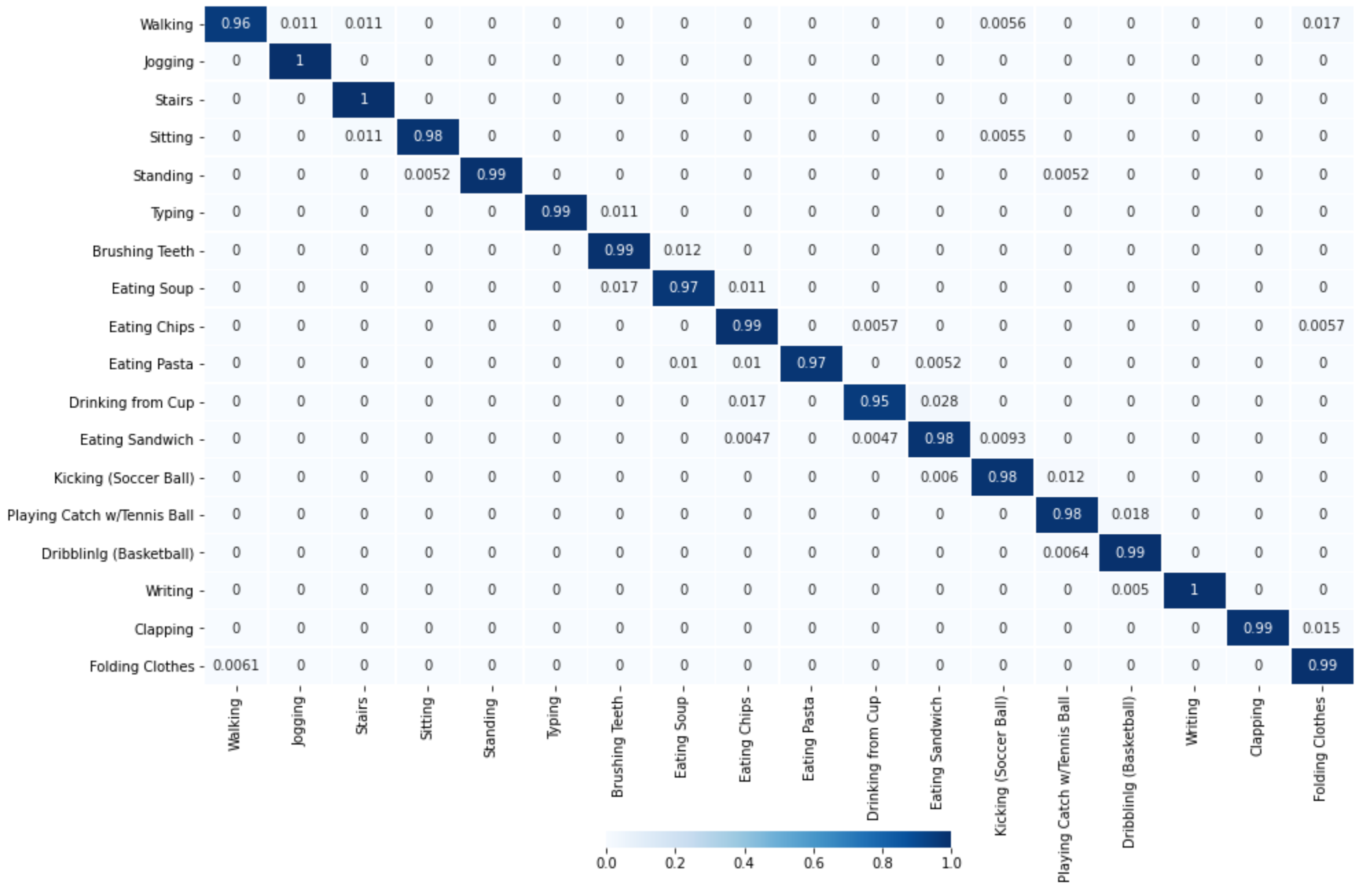


Fig. 8. Confusion matrix computed with 3-D acceleration (a_x^w, a_y^w, and a_z^w) using the WISDM 2019 dataset (smartwatch data).



	Walking	Jogging	Stairs	Sitting	Standing	Typing	Brushing Teeth	Eating Soup	Eating Chips	Eating Pasta	Drinking from Cup	Eating Sandwich	Kicking (Soccer Ball)	Playing Catch w/Tennis Ball	Dribblinlg (Basketball)	Writing	Clapping	Folding Clothes
Walking	0.96	0.011	0.011	0	0	0	0	0	0	0	0	0	0.0056	0	0	0	0	0.017
Jogging	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stairs	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sitting	0	0	0.011	0.98	0	0	0	0	0	0	0	0	0.0055	0	0	0	0	0
Standing	0	0	0	0.0052	0.99	0	0	0	0	0	0	0	0	0.0052	0	0	0	0
Typing	0	0	0	0	0	0.99	0.011	0	0	0	0	0	0	0	0	0	0	0
Brushing Teeth	0	0	0	0	0	0	0.99	0.012	0	0	0	0	0	0	0	0	0	0
Eating Soup	0	0	0	0	0	0	0.017	0.97	0.011	0	0	0	0	0	0	0	0	0
Eating Chips	0	0	0	0	0	0	0	0	0.99	0	0.0057	0	0	0	0	0	0	0.0057
Eating Pasta	0	0	0	0	0	0	0	0.01	0.01	0.97	0	0.0052	0	0	0	0	0	0
Drinking from Cup	0	0	0	0	0	0	0	0	0.017	0	0.95	0.028	0	0	0	0	0	0
Eating Sandwich	0	0	0	0	0	0	0	0	0.0047	0	0.0047	0.98	0.0093	0	0	0	0	0
Kicking (Soccer Ball)	0	0	0	0	0	0	0	0	0	0	0	0.006	0.98	0.012	0	0	0	0
Playing Catch w/Tennis Ball	0	0	0	0	0	0	0	0	0	0	0	0	0	0.98	0.018	0	0	0
Dribblinlg (Basketball)	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0064	0.99	0	0	0
Writing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.005	1	0	0
Clapping	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99	0.015
Folding Clothes	0.0061	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99

Fig. 9. Confusion matrices computed 6-D input signal (6-D acceleration and angular velocities $[a_x^w, a_y^w, a_z^w, \omega_x^w, \omega_y^w, \omega_z^w]$) using the WISDM 2019 dataset (smartwatch data).

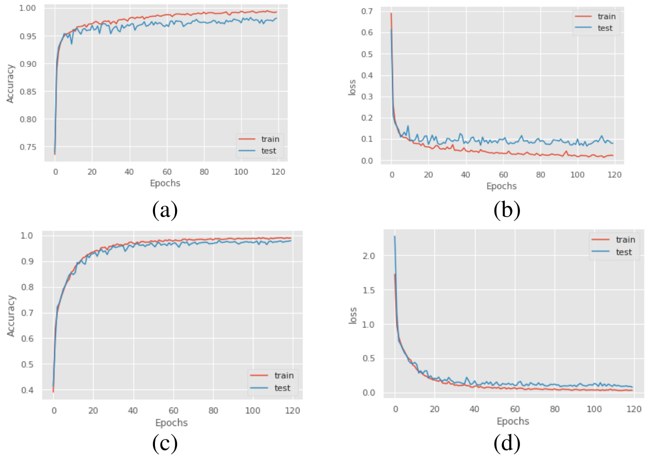
TABLE VI
NUMBERS OF TRAINABLE AND NONTRAINABLE PARAMETERS FOR DIFFERENT INPUT SIZES ARE PRESENTED

	WISDM 2011	
Cases	**Trainable Parameters**	**Non-trainable Parameters**
3D Acceleration	354,424	768
1D Magnitude of 3D Acceleration	354,048	768
	WISDM 2019	
6D Acceleration & Angular Velocities	361,144	768
3D Acceleration	360,580	768
3D Angular Velocities	360,580	768
1D Magnitude of 3D Angular Velocities	360,204	768
1D Magnitude of 3D Acceleration	360,204	768

the results that the bigger the window size, the better the performance. Moreover, the better the overlap between two consecutive segments, the better the classification performance of the model.

D. Accuracy and Loss Curves

Fig. 10 shows the accuracy and loss curves for training and testing that were obtained by using the magnitude of 3-D acceleration from smartphone accelerometers in the WISDM 2011 and WISDM 2019 datasets. The overlapping training and testing curves present a good fit without overfitting. When comparing the magnitude of 3-D acceleration from a smartwatch and smartphone accelerometers, similar behavior is seen.



(a)

(b)

(c)

(d)

Fig. 10. Accuracy and loss curves for WISDM 2011 and WISDM 2019 datasets ruling out overfitting. (a) Accuracy curves WISDM 2011 dataset. (b) Loss curves WISDM 2011 dataset. (c) Accuracy curves WISDM 2019 dataset. (d) Loss curves WISDM 2019 dataset.

E. Complexity of the Model

The aim of this study was to come up with a less complex model for HAR, which outperforms the state-of-the-art.

TABLE VII

WISDM 2011 Dataset—Comparison With the Existing Approaches

#	Reference	Year	Learner	Accuracy (%)	Recall/Sensitivity (%)	Specificity (%)
1	Peppas et al., [20]	2020	CNN + Statistical features	94.18	–	–
2	Xia et al., [22]	2020	LSTM-CNN	95.85	–	–
3	Li et al., [24]	2020	CNN-LSTM (after feature enhancement)	99.47	–	–
4	Imran and Latif [9]	2020	CNN	95.26	95	–
5	Mehmood et al., [10]	2020	CNN	94.65	95	–
6	Zhang et al., [19]	2020	CNN with attention mechanism	96.4	–	–
7	Dua et al., [23]	2021	CNN+RF	97.77	–	–
8	Nafea et al., [25]	2021	CNN+BiLSTM	98.53	–	–
9	Imran et al., [26]	2022	BiGRU+CNN	97.20	97	–
10	Imran et al., [29]	2022	CNN	96.62	97	–
11	Imran et al., [30]	2022	CNN	89.60	90	–
12	Proposed Approach	2023	CNN-BiGRU (using a_x^p, a_y^p, a_z^p)	98.81	99	97.84
13	Proposed Approach	2023	CNN-BiGRU (using $\widehat{mag_a^p}$)	97.29	97	99.08

TABLE VIII

WISDM 2019 Dataset—Comparison With the Existing Approaches

#	Reference	Year	Learner	Features	Accuracy (%)	Recall/Sensitivity (%)	Specificity (%)
1	Ihianle et al., [31]	2020	MCBLSTM	$(a_x^w, a_y^w, a_z^w, \omega_x^w, \omega_y^w, \omega_z^w)$	96.6±1.47	–	–
2	Weiss et al., [13]	2019	KNN, DT, RF	$(a_x^w, a_y^w, a_z^w, \omega_x^w, \omega_y^w, \omega_z^w)$ + $(a_x^p, a_y^p, a_z^p, \omega_x^p, \omega_y^p, \omega_z^p)$	94.4	–	–
3	Proposed Approach	2023	CNN-BiGRU	$a_x^w, a_y^w, a_z^w, \omega_x^w, \omega_y^w, \omega_z^w$	98.4	98	97.61
4				a_x^w, a_y^w, a_z^w	98.2	98	98.50
5				$\widehat{mag_a^w}$	97.5	98	99.30
6				$\omega_x^w, \omega_y^w, \omega_z^w$	96.8	97	97.33
7				$\widehat{mag_\omega^w}$	63.9	67	89.24

We experimented with different cases of smartwatches and smartphone IMU sensors. It has been shown that the use of the magnitude of an accelerometer produces comparable performance. Regarding trainable and nontrainable parameters for both datasets, Table VI presents a comparison of the proposed model's complexity on various input sizes.

VI. Conclusion

This research presents a novel architecture that is capable of classifying 18 daily life activities with higher classification accuracies. We also propose the magnitude of 3-D acceleration ($\widehat{\text{mag}_a}$) to convert the 3-D input signal into 1-D, thereby significantly reducing the computation cost. For the sake of comparison, we also consider 3-D input (3-D acceleration and 3-D angular velocities) and 6-D input (triaxial acceleration and triaxial angular velocities). We have tested the proposed model on two well-known HAR datasets, i.e., WISDM 2011 (data collected with a smartphone under a controlled environment) and WISDM 2019 (data collected with a smartwatch in an open environment without any environmental constraints/controlling). With the 1-D input signal, i.e., the magnitude of 3-D acceleration ($\widehat{\text{mag}_a}$), we achieved an average classification accuracy of 97.5%, whereas, with the 3-D and 6-D input signal, the average classification accuracy is improved by 1% only. The presented work is also compared with existing works and it not only outperformed them (regarding accuracy) but also reduced the complexity of the problem by making use of fewer features for inference.

Given the higher classification accuracy and the model being lightweight (due to the 1-D input signal as well as the model's architecture), we believe that the proposed model paves the path to real-time HAR on edge devices such as smartwatches and smart fitness bands. Since the validation is done on a smartwatch and we have considered a range of locomotive and nonlocomotive activities, a real-time health monitoring application can greatly help the general public to keep track of the type of daily activities they are involved in. This can lead to making better decisions in order to promote healthy lifestyles and well-being of the individuals.

In terms of limitations of the work, the proposed model has over 0.7 million parameters, which can be minimized. Further reduction in the trainable parameters is an important future direction so that the proposed approach can be deployed on low-end, low-energy edge devices. In a similar direction, one can think of employing an attention mechanism, which is known to be more efficient than CNNs and RNNs. Similarly, a large bunch of activities of daily living (locomotive as well as nonlocomotive) should be included in the training and validation so that the deep model can classify a wide range of activities of daily living.

A. Comparison With Current Studies

A comprehensive comparison with the existing HAR methodologies that were trained and tested over the WISDM 2011 and 2019 datasets has been carried out. Most of the existing studies used WISDM 2011 dataset and we were able to find only a few works that explored the WISDM 2019 dataset. Hence, the newer dataset, which is collected under a more realistic and open environment using a smartwatch, is not extensively analyzed and explored by the research community. The comparison results of WISDM 2011 dataset are summarized in Table VII, whereas Table VIII presents a comparison with the existing approach for WISDM 2019 dataset.

Our proposed work is compared with a total of ten recently published architectures for HAR using the WISDM 2011 dataset. The presented model outperforms the existing methodologies regarding classification accuracy. With 3-D acceleration as input, we achieved an average classification accuracy of 98.81%. Furthermore, with the $\widehat{\text{mag}}_a^w$ (1-D input), the average classification accuracy only drops slightly and remains at 97.29%. The computation cost of the $\widehat{\text{mag}}_a^w$, being a 1-D input signal, is much lower than that of a higher dimensional input signal (as used in most of the existing studies).

For the WISDM 2019 dataset, the number of published studies is very limited. During the comparison, we found that our proposed model outperforms the study presented by Weiss et al. [13] regarding classification accuracies. The deep model proposed by Ihianle et al. [31] relies on CNN-BiLSTM and they reported an accuracy of 96.6 ± 1.47% when the model is trained with the smartwatch data. The input signal consists of 6-D acceleration and angular velocities making the model computationally expensive. In comparison, we have achieved an average classification accuracy of 98.4% with 6-D input size (6-D acceleration and angular velocities), 98.2% with 3-D acceleration, 96.8% with 3-D angular velocities, and 97.5% with 1-D input signal, i.e., magnitude of accelerations (see Table VIII for details). The latter case, being a 1-D input signal, is computationally efficient as it requires minimal computation cost. Thus, it is more feasible when it comes to implementing it in the real world.

References

[1] J. Lu, X. Zheng, M. Sheng, J. Jin, and S. Yu, “Efficient human activity recognition using a single wearable sensor,” *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11137–11146, Nov. 2020.

[2] R. Desai, N. E. Fritz, L. Muratori, J. M. Hausdorff, M. Busse, and L. Quinn, “Evaluation of gait initiation using inertial sensors in Huntington’s disease: Insights into anticipatory postural adjustments and cognitive interference,” *Gait Posture*, vol. 87, pp. 117–122, Jun. 2021.

[3] H. A. Imran, Q. Riaz, M. Zeeshan, M. Hussain, and R. Arshad, “Machines perceive emotions: Identifying affective states from human gait using on-body smart devices,” *Appl. Sci.*, vol. 13, no. 8, p. 4728, Apr. 2023.

[4] M. A. Hashmi, Q. Riaz, M. Zeeshan, M. Shahzad, and M. M. Fraz, “Motion reveal emotions: Identifying emotions from human walk using chest mounted smartphone,” *IEEE Sensors J.*, vol. 20, no. 22, pp. 13511–13522, Nov. 2020.

[5] V. B. Semwal, N. Gaud, P. Lalwani, V. Bijalwan, and A. K. Alok, “Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor,” *Artif. Intell. Rev.*, vol. 55, no. 2, pp. 1149–1169, Feb. 2022.

[6] I. Gohar, Q. Riaz, M. Shahzad, M. Z. U. H. Hashmi, H. Tahir, and M. E. U. Haq, “Person re-identification using deep modeling of temporally correlated inertial motion patterns,” *Sensors*, vol. 20, no. 3, p. 949, Feb. 2020.

[7] P. Bharti, D. De, S. Chellappan, and S. K. Das, “HuMAn: Complex activity recognition with multi-modal multi-positional body sensing,” *IEEE Trans. Mobile Comput.*, vol. 18, no. 4, pp. 857–870, Apr. 2019.

[8] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, “Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey,” *IEEE Access*, vol. 8, pp. 210816–210836, 2020.

[9] H. A. Imran and U. Latif, “HHARNet: Taking inspiration from inception and dense networks for human activity recognition using inertial sensors,” in *Proc. IEEE 17th Int. Conf. Smart Communities, Improving Quality Life Using ICT, IoT AI (HONET)*, Dec. 2020, pp. 24–27.

[10] K. Mehmood, H. A. Imran, and U. Latif, “HARDenseNet: A 1D DenseNet inspired convolutional neural network for human activity recognition with inertial sensors,” in *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, Nov. 2020, pp. 1–6.

[11] X. Chen, S. Jiang, and B. Lo, “Subject-independent slow fall detection with wearable sensors via deep learning,” in *Proc. IEEE Sensors*, Nov. 2020, pp. 1–4.

[12] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, “Design considerations for the WISDM smart phone-based sensor mining architecture,” in *Proc. 5th Int. Workshop Knowl. Discovery Sensor Data*, Aug. 2011, pp. 25–33.

[13] 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.

[14] D. Anguita et al., “A public domain dataset for human activity recognition using smartphones,” in *Proc. ESANN*, vol. 3, 2013, p. 3.

[15] P. Foudeh, A. Khorshidtalab, and N. Salim, “A probabilistic data-driven method for human activity recognition,” *J. Ambient Intell. Smart Environ.*, vol. 10, no. 5, pp. 393–408, Sep. 2018.

[16] O. Banos et al., “mHealthDroid: A novel framework for agile development of mobile health applications,” in *Proc. Int. Workshop Ambient Assist. Living*. Cham, Switzerland: Springer, 2014, pp. 91–98.

[17] V. Soni, H. Yadav, V. B. Semwal, B. Roy, D. K. Choubey, and D. K. Mallick, “A novel smartphone-based human activity recognition using deep learning in health care,” in *Machine Learning, Image Processing, Network Security and Data Sciences*. Cham, Switzerland: Springer, 2023, pp. 493–503.

[18] S. K. Challa, A. Kumar, and V. B. Semwal, “A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data,” *Vis. Comput.*, vol. 38, no. 12, pp. 4095–4109, Dec. 2022.

[19] H. Zhang, Z. Xiao, J. Wang, F. Li, and E. Szczerbicki, “A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention,” *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1072–1080, Feb. 2020.

[20] K. Peppas, A. C. Tsolakis, S. Krinidis, and D. Tzovaras, “Real-time physical activity recognition on smart mobile devices using convolutional neural networks,” *Appl. Sci.*, vol. 10, no. 23, p. 8482, Nov. 2020.

[21] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SIGKDD Explor. Newslett.*, vol. 12, no. 2, pp. 74–82, Mar. 2011.

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

[23] N. Dua, S. N. Singh, and V. B. Semwal, “Multi-input CNN-GRU based human activity recognition using wearable sensors,” *Computing*, vol. 103, no. 7, pp. 1461–1478, Jul. 2021.

[24] X. Li, L. Nie, X. Si, R. Ding, and D. Zhan, “Enhancing representation of deep features for sensor-based activity recognition,” *Mobile Netw. Appl.*, vol. 26, no. 1, pp. 130–145, Feb. 2021.

[25] 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.

[26] H. A. Imran, “UltaNet: An antithesis neural network for recognizing human activity using inertial sensors signals,” *IEEE Sensors Lett.*, vol. 6, no. 1, pp. 1–4, Jan. 2022.

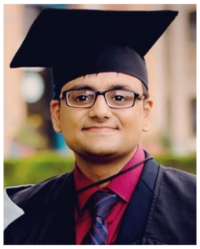
[27] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K.-R. Choo, “Imaging and fusing time series for wearable sensor-based human activity recognition,” *Inf. Fusion*, vol. 53, pp. 80–87, Jan. 2020.

[28] Q. Riaz, G. Tao, B. Krüger, and A. Weber, “Motion reconstruction using very few accelerometers and ground contacts,” *Graph. Models*, vol. 79, pp. 23–38, May 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1524070315000107

[29] H. A. Imran, K. Hamza, and Z. Mehmood, “HARResNext: An efficient ResNext inspired network for human activity recognition with inertial sensors,” in *Proc. 2nd Int. Conf. Digit. Futures Transformative Technol. (ICoDT2)*, May 2022, pp. 1–4.

[30] H. A. Imran, “Khail-Net: A shallow convolutional neural network for recognizing sports activities using wearable inertial sensors,” *IEEE Sensors Lett.*, vol. 6, no. 9, pp. 1–4, Sep. 2022.

[31] I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa, and M. O. Orisatoki, “A deep learning approach for human activities recognition from multimodal sensing devices,” *IEEE Access*, vol. 8, pp. 179028–179038, 2020.



Hamza Ali Imran received the B.S. degree in electrical engineering from the National University of Computer and Emerging Science (NUCES-FAST), Islamabad, Pakistan, in 2018, and the M.S. degree in computer science from the School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad, in 2022.

He has over more than four years of industrial experience in the field of embedded systems and the Internet of Things. Currently, he is serving at the ML Group, Emumba Pvt. Ltd., Islamabad. He has also worked as a Research Assistant at the Embedded Systems and Pervasive Computing (EPIC) Laboratory, NUCES-FAST, where he focused on high-performance computing. His research interests include embedded systems, embedded sensors, system engineering, human motion analysis, and deep learning.

Qaiser Riaz (Senior Member, IEEE) received the M.S. degree in autonomous systems from the Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany, in 2011, and the Ph.D. (Dr. rer. nat.) degree in computer science from the University of Bonn, Bonn, Germany, in 2016.

He is currently working as a tenured Associate Professor at the Department of Computing, School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan. His research interests include human motion analysis and synthesis using low-cost wearable sensors, wearable inertial measurement units (IMUs) for smart healthcare, the Internet of Things (IoT), deep learning, character animation, and network security.

Mehdi Hussain (Senior Member, IEEE) received the B.S. degree in computer science from Islamia University Bahawalpur, Bahawalpur, Pakistan, in 2006, the M.S. degree in computer science from the Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi, Pakistan, in 2011, and the Ph.D. degree from the University of Malaya, Kuala Lumpur, Malaysia, in 2017.

From 2006 to 2014, he worked in various software engineering positions in renowned U.S.-based software houses in Pakistan. He is currently working as an Assistant Professor at the School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad. He has experience with various video/image-based projects (i.e., Video Conferencing, H.264 Encoding, and License Plate Recognition). His primary area of research is information security, with a particular focus on information hiding, digital forensics, and multimedia security.

Hasan Tahir (Senior Member, IEEE) received the M.S. degree in software engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2011, and the Ph.D. degree in computing and electronics from the University of Essex, Colchester, U.K., in 2017.

He has taught several courses at UG and PG levels. His areas of interest include but are not limited to physical unclonable function (PUF), applied cryptography, network security, the Internet of Things (IoT) security, and cloud computing.

Razi Arshad (Senior Member, IEEE) received the M.Sc. degree in information security from Sichuan University, Chengdu, China, in 2007, and the Ph.D. degree in mathematical cryptography (with a focus on symmetric key cryptography) from Otto-von-Guericke University, Magdeburg, Germany, in 2018.

He is a Research Fellow in Security and Privacy at the University of Surrey, Guilford, U.K. Prior to this position at the University of Surrey, he was working as an Assistant professor at the School of Electrical Engineering and Computer Sciences, National University of Sciences and Technology, Islamabad, Pakistan.