IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 October 2023, accepted 19 November 2023, date of publication 30 November 2023,
date of current version 8 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3338137

RESEARCH ARTICLE

Attention Mechanism-Based Bidirectional Long Short-Term Memory for Cycling Activity Recognition Using Smartphones

VAN SY NGUYEN[1,2], HYUNSEOK KIM[3], AND DONGJUN SUH[1], (Member, IEEE)
[1]Department of Convergence and Fusion System Engineering, Kyungpook National University, Sangju 37224, South Korea
[2]Mechanical and Aerospace Engineering Department, University of Central Florida, Orlando, FL 32816, USA
[3]Department of Computer Engineering, Dong-A University, Busan 49315, South Korea

Corresponding author: Dongjun Suh (dongjunsuh@knu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant NRF-2021R1A5A8033165, and in part by the Basic Science Research Program through the NRF funded by the Ministry of Education under Grant NRF-2021R1I1A3049503.

ABSTRACT Bicycles are an ecofriendly mode of transportation, and cycling offers physical and mental well-being. However, their increased use has resulted in frequent bicycle–human accidents, car-to-bicycle collisions, related injuries and cyclist crashes. Moreover, rules for safe cycling are limited. Smart healthcare systems using smartphones and/or wearable devices, such as a cycling monitoring application that can inform fellow cyclists about the state of the user, can be developed to provide assistance during such unexpected events. In this study, a one-dimensional convolutional neural network (1DCNN)–bidirectional long short-term memory (BiLSTM) based on attention mechanism (CBiAM) model is proposed for detecting cyclists' states using a mobile phone, thereby enhancing their safety and promoting a secure cycling experience in case of accidents or emergencies. In addition, the ''cycling safe (CySa) dataset,'' a new dataset containing data on the cyclists' actions during cycling, collected from a smartphone positioned in the cyclists' pocket is presented. The proposed CBiAM model was trained on the CySa dataset using different sliding window sizes, batch sizes (Bz), and learning rates (Lr). Experimental results confirmed the superior performance of the proposed model compared to conventional approaches, such as support vector machines and artificial neural networks, and existing advanced architectures, such as 1DCNN, long short-term memory (LSTM), and Bi-LSTM. The robustness of the model was validated using public datasets, such as UCI-human activity recognition (HAR), PAMAP2, Opportunity, MOTIONSENSE, and WISDM, where it achieved impressive F1-scores of 97.51%, 99.82%, 94.72%, 97.67%, and 87.05%, respectively.

INDEX TERMS Attention mechanism, cycling, human activity recognition, bidirectional LSTM.

I. INTRODUCTION

Cycling has become increasingly popular as an excellent form of physical exercise that can improve cardiovascular health, strengthen muscles, and boost mental well-being [1]. Cycling is also cost effective as it does not incur fuel costs and insurance or parking fees and is free of harmful emissions [2]. However, it involves some potential risks. Bicycles are a common means of transportation, and in a study conducted in a large city in Germany, the number of bicycle accidents accounted for almost half of all the traffic accidents in the city [3]. In 2010–2014, approximately 32% of cyclists in Sweden, suffered from serious injuries due to bicycle accidents [4]. Therefore, the condition of fellow cyclists should be accessible to offer quick support in case of bicycle crashes or emergencies, such as administering first aid, calling for emergency services, or transporting the injured person to a hospital.

At present, mobile devices (off-body sensing) and/or wearable devices (on-body sensing) are commonly used for monitoring human activities [5]. Smartphones (combined with off-body and on-body sensing) are equipped with a

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

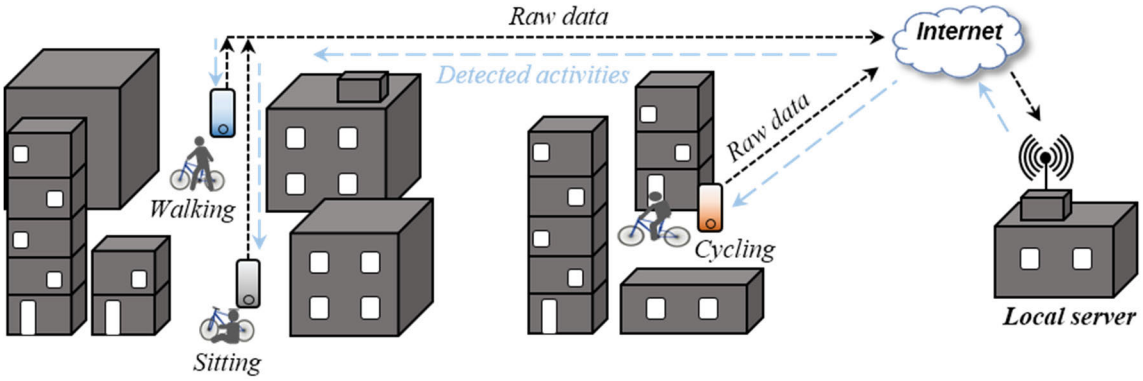

FIGURE 1. Block diagram of a bicycle assistance application.

range of sensing applications that can measure various health-related data, such as heart rate, blood pressure, and oxygen saturation [6]. Mobile apps collect sensor data and send it to a server for analysis. Health care professionals can leverage these datasets to remotely evaluate their patients' conditions and adjust their medication dosages by sending them messages based on the analyzed data.

Similarly, mobile devices can be used in cycling groups for improved assistance in case of emergencies. Fig. 1 shows a block diagram of a bicycle assistance application, where a mobile with various sensors can capture data from each cyclist in the group. The raw data is then transmitted to a server via the Internet (cloud infrastructure) or SMS [7]. Using artificial intelligence, the data can be analyzed and processed by a deep learning (DL) model to predict the condition of each cyclist. Thus, cyclists can have real-time access to the condition of all members during cycling and offer prompt assistance when needed. Additionally, with advanced processing capabilities and the development of mobile applications, the DL model can be integrated into mobile devices to perform classification tasks and conveniently transmit prediction results to other devices.

Furthermore, with easy accessibility of mobile and wearable devices due to their high energy efficiency, compactness, and low cost, multiple smartphone built-in sensors can now be utilized for activity recognition. Accelerometers, gyroscopes, and magnetometers are widely used sensors on mobile devices for activity recognition owing to their popularity and ease of data collection [8]. An accelerometer measures a mobile device's acceleration in three directions (x, y, and z) to detect movements such as walking, running, and jumping. A gyroscope measures the device's rotation around these three axes to evaluate activities such as turning or falling. Lastly, a magnetometer measures the Earth's magnetic field in three dimensions to determine the device orientation, which helps recognize activities such as standing or lying down. Thus, a cyclist's activity can be effectively analyzed by combining the information obtained from these three sensors.

Moreover, positions can be tracked using mobile phones equipped with the global positioning system (GPS), WiFi, or location-based services (LBSs). GPS can achieve an accuracy of 9 and 15 m for horizontal and vertical positioning, respectively [9]. And the position of fast-moving objects such as bicycles can be easily detected using GPS sensor data. In addition, the transmission of data between a mobile phone and local server is realistic and feasible nowadays. Thus, a DL model must be developed for predicting the movement of cyclists to ensure their safety.

The remainder of this paper is organized as follows: Section II presents an overview of current research. Section III outlines the dataset collection, feature selection, and feature preprocessing, and the proposed one-dimensional convolutional neural network (1DCNN)–bidirectional long short-term memory (BiLSTM) based on attention mechanism (CBiAM) architecture is compared with traditional methods. Section IV presents the experimental results and discussion. Section V concludes the study.

II. RELATED WORKS

In the past few years, several methods have been proposed for modeling and recognizing human activities. To our knowledge, although cycling activities have been studied, human activity recognition (HAR) specifically designed to assist cycling groups has not been reported yet. For instance, with an aim to monitor activity using smartphone sensors and improve lifestyle, Mannini [10] developed accelerometer-based activity monitors to classify four activities, namely ambulation, cycling, sedentary, and painting, using a support vector machine (SVM) and achieved an overall classification accuracy of 84.7%. To improve the classification accuracy, gyroscope and accelerometer data were also used to train several classifiers, including Naive Bayes, decision tree, k-nearest neighbors (KNN), and SVM [11]. The decision tree classifier was identified as the best model because it distinguished various actions with high accuracy, such as climbing stairs, descending stairs, driving, and cycling. A wristband with heart rate and accelerator sensors

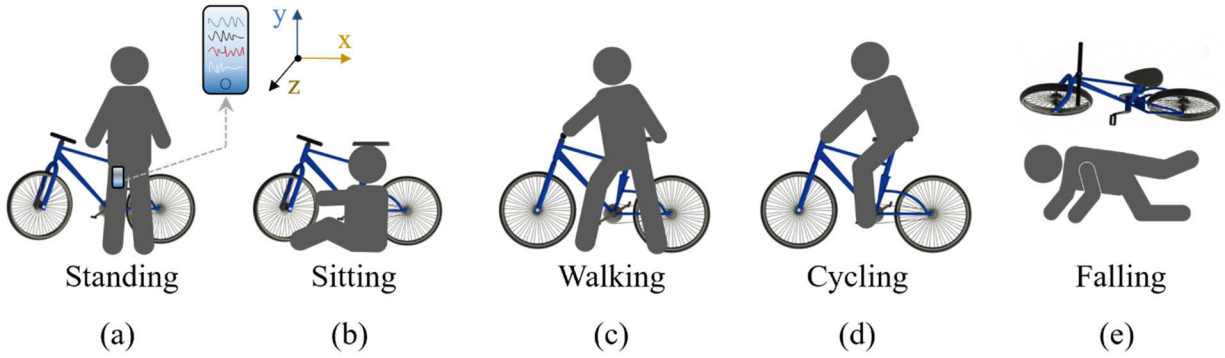

FIGURE 2. Placement of smartphones on participants and the body gestures corresponding to each action.

was used to detect specific activities at home [12]. Therein, four basic activities, such as standing, sitting, householding, and cycling with two intensity levels, were considered, and random forest (RF) and SVM classifiers were applied on the collected data.

For RF and SVM classifiers, the average recognition accuracies with leave-one-subject-out cross-validation were 89.2% and 85.6%, respectively, and F1-score values were 0.89 and 0.86, respectively. A study was conducted on daily activity classifiers for five common activities, such as cycling, running, sitting, standing, and walking, for health care applications using acceleration and gyroscope sensors to manage chronic diseases. The AdaBoost algorithm demonstrated good performance, achieving a classification accuracy of 98% [13].

DL has garnered attention from developers in HAR. Several common DL architectures, such as artificial neural networks (ANNs) [14], [15], one-dimensional convolutional neural networks (1DCNNs), auto-encoders, long short-term memory (LSTM), bidirectional LSTM, and 1DCNN–LSTM provide high performance in complex classification tasks and can be easily employed in mobile apps, making them feasible for further development and wide usage. For example, an ANN-based classifier was designed to identify user identity or find the owner of a phone based on the acceleration data of walking recorded while the phone was in the user's pocket [15]. The multilayer perceptron (MLP) achieved 98.11% accuracy in gesture recognition but required more processing time than SVM.

Furthermore, 1DCNNs and LSTM models have gained popularity due to their remarkable performance in various classification problems [16], [17], [18], [19]. For instance, a 1DCNN-based model was employed for pedestrian activity recognition [16], which achieved approximately 98% accuracy in classifying nine types of activities within 2 s. Additionally, a HAR classifier based on a 1DCNN model [17] predicted two similar actions (going upstairs and walking) at an accuracy of 96.11%.

Moreover, LSTM-based classifiers have shown outstanding ability in handling sequences with long-term dependencies. For example, a model was designed to identify everyday activities of individuals with COVID-19 and provide appropriate recommendations [18], which yielded a higher classification accuracy of 97.33% compared to traditional approaches such as decision trees, SVM, and KNN. Similarly, LSTM-based models have been used for HAR to improve the quality of life of residents [19] which obtained higher classification accuracies compared to traditional approaches such as Naive Bayes or conditional random fields.

Additionally, a multikernel block CNN-based classifier was proposed to predict stationary and nonstationary activities (such as cycling and walking). The model exhibited a high classification accuracy compared to data fusion techniques, RF, and CNN-RNN models [20]. Similarly, an ensemble model (CNN-LSTM) was used for human activity recognition to evaluate various activities on different public datasets [21], such as UCI-HAR, WISDM, OPPORUNITY, which achieved a high classification accuracy of 95.78%, 95.85%, and 92.63%, respectively.

The use of Convolutional Neural Networks (CNN) and/or Gated Recurrent Units (GRU) is an effective solution for recognizing human activities. For example, generating images from sensor readings and using CNN to classify human activities has shown promising results and has enhanced classification outcomes [22], [23]. The combination of CNN and GRU, based on the Inception architecture, can capture not only local patterns at multiple scales but also long-term dependencies in the sequence data, providing robust detection in activity classification [24].

Attention mechanism (AM) has recently been applied and shown promising results in several applications [25], [26], [27], [28]. An AM-based multihead model was proposed for HAR, which achieved higher accuracy compared to traditional techniques. Yin et al. conducted a study on the combination of CNN-RNN and AM to enhance the classification performance. The classification accuracy using this combination was higher than that when only CNN or RNN was used [28].

In this study, a cycling safe (CySa) dataset comprising data on five basic activities, such as standing, sitting, walking with a bicycle, cycling, and falling, was developed.

Additionally, an AM-based CNN-Bidirectional LSTM (CBiAM) model was proposed and applied to improve classification accuracy. Hyperparameter adjustment was performed to assess the impact of hyperparameters on the model performance. The proposed model was compared with other classification methods, and the results confirmed that the proposed model achieved high accuracy compared to the existing architectures. Finally, the robustness of the proposed model on public datasets, such as UCI-HAR, PAMAP2, Opportunity, MOTIONSENSE, and WISDM, was demonstrated.

III. MATERIAL AND METHODS

A. DATASET DESCRIPTION AND PREPROCESSING

1) DATASET COLLECTION

CySa was created by collecting data from mobile sensors (sensor logger on iPhone 8). Five participants (three males and two females of 25–30 years) performed five actions: standing, sitting, walking with a bicycle, cycling, and falling from a bicycle. Each action was performed for approximately 3 min with a mobile phone placed in the participant's front pocket to capture the accelerometer (A_x, A_y, A_z), gyroscope (G_x, G_y, G_z), and magnetometer (M_x, M_y, M_z) data, as shown in Fig. 2. To simulate cycling accidents, the falling action was defined as when a person falls off the bicycle and rotates several times to the side. The falling action was designed to be quick to distinguish it from simply getting off the bicycle. At a sampling rate of 10 Hz [29], each action performed by one person contained around 1800–2000 data points, accounting to around 10,000 data points in total for all five participants. The total number of data points in the CySa dataset is approximately 50,000 samples, as shown in Table 1.

TABLE 1. Number of data points of the cycling safe dataset[1].

Activities	Abbreviations	Samples
Standing	Stan	10902
Sitting	Sit	11472
Walking	Wal	10256
Cycling	Cyc	11334
Falling	Fal	10789

2) FEATURE EXTRACTION SELECTION

a: FEATURE SELECTION

After the CySa dataset was collected, different features from that raw dataset were extracted, such as the mean (Mea), median (Med), variance (Va), and standard deviation (St), which are the most used features in HAR [30]. Mea, Med, Va and St of features were calculated using (1)–(4), respectively. Specifically, Mea^i_m is the median value with a window size (k), where i is the number of nonoverlapping windows or is the number of extracted features created by each sensor, as shown in Fig. 3. m refers to the type of sensor values. As 9 sensor values (A_x, A_y, A_z; G_x, G_y, G_z; and M_x, M_y, M_z) were used, m ranges from 1 to 9.

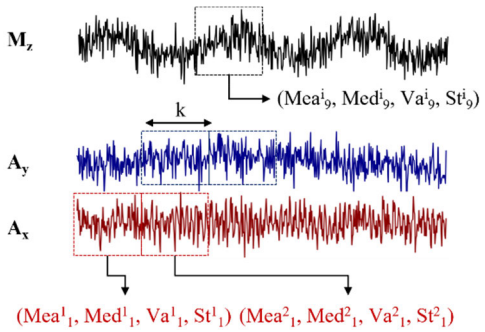

FIGURE 3. Feature selection of the CySa dataset.

The length of the samples (k) has a significant effect on both the computation time and the ability to identify activity details. The selection of appropriate sample length varies according to applications, the sample length of 1–5s is recommended for complex activity recognition [8]. Herein, k was chosen as 10 and a nonoverlapping procedure was used for feature selection [31].

$$Mea^i_{A_x} = \frac{1}{k}\sum_{i=1}^{k}(A_x)_i\,, \tag{1}$$

$$Med^i_{A_x} = \left((A_x)_{\frac{k}{2}} + (A_x)_{\frac{k}{2}+1}\right)/2, \tag{2}$$

$$Var^i_{A_x} = \frac{\sum_{i=1}^{k}\left((A_x)_i - Mea^i_{A_x}\right)^2}{k}, \tag{3}$$

$$St^i_{A_x} = \sqrt{\frac{\sum_{i=1}^{k}\left((A_x)_i - Mea^i_{A_x}\right)^2}{k}}. \tag{4}$$

where Mea^i_{Ax}, Med^i_{Ax}, Var^i_{Ax} and St^i_{Ax} are mean, median, variance, and standard deviation of the accelerometer sensor in x-direction, which is calculated by a window size of k and a nonoverlapping window of i.

b: FEATURE NORMALIZATION

As the range of the dataset varies, the input must be normalized to avoid training bias. Herein, the datasets are mapped to a new range and min-max normalization is used for the input data. The equation is shown below:

$$x^n(i) = \frac{x^o(i) - x^o(i)_{min}}{x^o(i)_{max} - x^o(i)_{min}}\left(x^n_{max} - x^n_{min}\right) + x^n_{min} \tag{5}$$

where $x^n(i)$ is the normalized feature value of i-th sample, $x^o(i)_{max}$ and $x^o(i)_{min}$ are the maximum and minimum feature values in the original dataset, and $x^n_{max} = 1$ and $x^n_{min} = 0$ are the maximum and minimum feature values of the new range.

c: DATASET VISUALIZATION

Figs. 4(a)–(e) show the magnitudes of the accelerometer, gyroscope, and magnetometer in the x, y, and z directions during cycling, falling, sitting, standing, and walking, respectively. Cycling, falling, and walking produced large oscillations in the accelerometer, gyroscope, and magnetometer, as shown in Figs. 4 (a), (b), and (e). Understandably, these actions comprise a series of alternating leg movements, which can result in sudden changes and vibrations

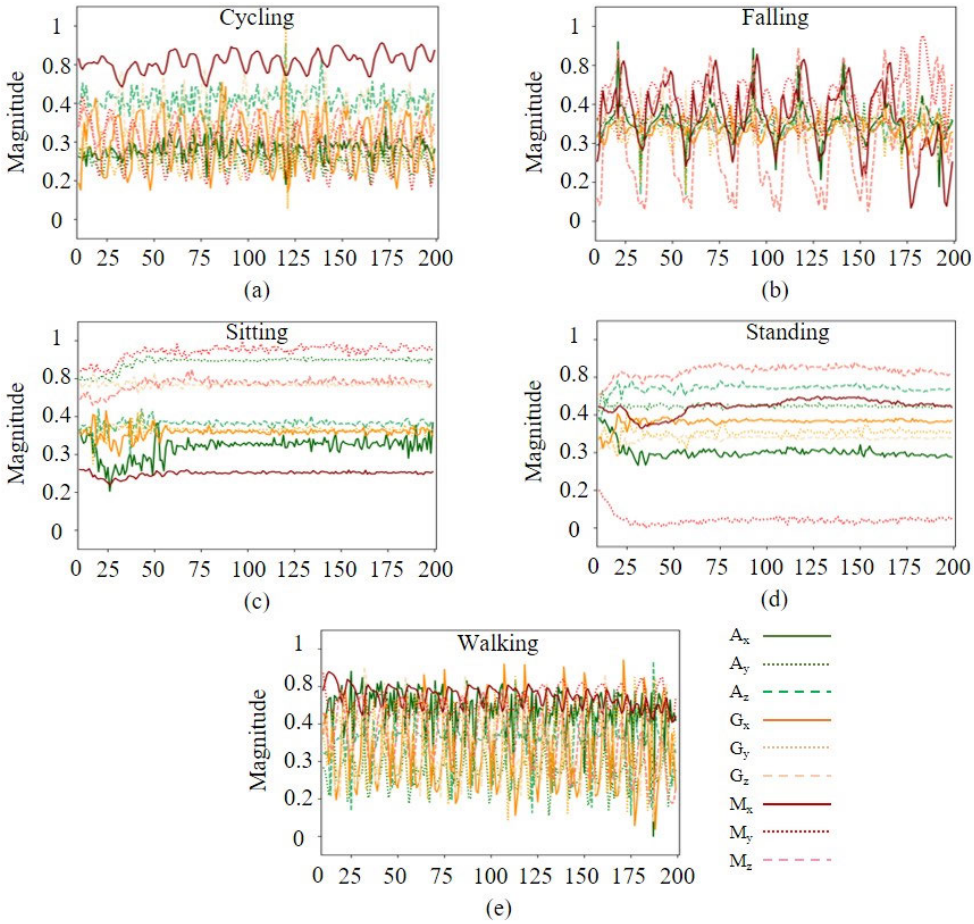

FIGURE 4. Normalized magnitude of the sensor values. (a) Cycling, (b) falling, (c) sitting, (d) standing, and (e) walking.

and significant oscillations in the sensor readings. On the other hand, as the legs are in stable positions with minimal movement while sitting and standing, they result in small oscillations in the accelerometer and gyroscope readings, as shown in Figs. 4(c) and (d).

B. NEURAL NETWORK ARCHITECTURE OF THE PROPOSED MODEL

This section discusses several neural network architectures used to construct the proposed model. The descriptions of these architectures are as follows:

1) 1DCNN ARCHITECTURE

The 1DCNN model is a variant of the 2DCNN model; it has been developed and applied for analyzing sequential data, such as audio signals, or natural language processing tasks [32]. It is composed of an input layer, convolutional layers, pooling layers, and fully connected layers. The input of 1D-CNN is (M × N) matrix, where M is the length of the data (steps or a sequence of data) and N is the channel (number of features). The kernel size has the shape of K × N, where K is given by a designer. For example, as shown in Fig. 5, K = 2, so the kernel size is the size of a 2 × 2 matrix. It will strike from left to right to create convolution maps. There are two main types of convolutions: overlapping and nonoverlapping. Overlapping convolution, which is used in the proposed model, can be useful when a high-resolution output has to be generated. Nonoverlapping convolution is typically used to reduce the computational complexity of convolution operations.

A pooling layer is a new layer added after the convolutional layer, and it slides over all the regions of the convolution map according to its stride. The pooling operation reduces the number of parameters and computation. Max and average pooling layers are commonly used in CNNs.

In a 1DCNN model, the convolutional layers extract feature maps by convolving the input signals with learned filters. These feature maps are then flattened into a 1D array and then fed into a flattened layer. Finally, a fully connected layer is used to produce outputs for classification or regression tasks.

2) 1DCNN-BILSTM ARCHITECTURE

The CNN–BiLSTM architecture combines the 1DCNN and BiLSTM models. The integration of 1DCNN and LSTM/BiLSTM networks has proven particularly effective in handling temporal patterns and 1D data. Using this

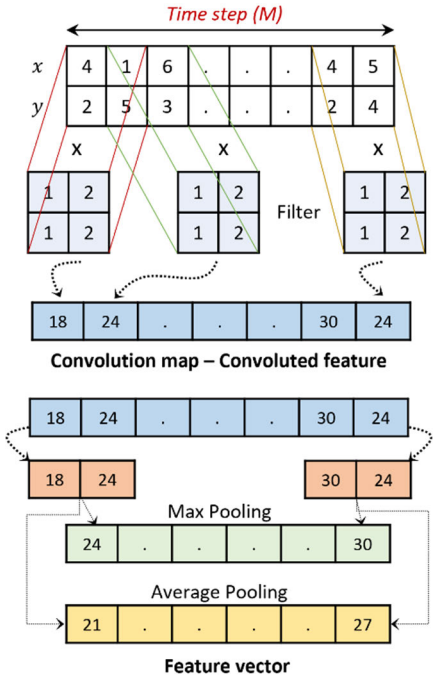

FIGURE 5. Architecture of the 1DCNN model.

combination, time-dependent features can be captured and information loss resulting from convolutional operations can be mitigated. As shown in Fig. 6(a), the convolutional layers, pooling layers, and fully connected layers in the 1DCNN model extract spatial features from the input data. Then, the valuable features go into the LSTM or BiLSTM layers. The BiLSTM model is a variant of the LSTM model [33], but it can process the sequence in both forward and backward directions, as shown in Fig. 6(c). As depicted in Fig. 6(b), each cell of the BiLSTM model comprises several main components such as input (i), output (o_t), and forget (f_t) gates. The i gate uses the current input (x_t) and previous hidden state (h_{t-1}) to determine the length of the incoming information. The o_t gate uses the current input and the previous hidden state and then passes them through a sigmoid activation function. Finally, the f_t gate decides whether to retain or omit the information from the previous state (C_{t-1}). The mathematical formulas for these gates are shown in (6)–(10):

$$i_t = sig\left(W_i.\left[h_{t-1}, X_t\right] + b_i\right), \tag{6}$$

$$f_t = sig\left(W_f.\left[h_{t-1}, X_t\right] + b_f\right), \tag{7}$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes tanh\left(W_C.\left[h_{t-1}, X_t\right] + b_C\right), \tag{8}$$

$$o_t = sig\left(W_o.\left[h_{t-1}, X_t\right] + b_o\right), \tag{9}$$

$$h_t = o_t \otimes \tanh(C_t), \tag{10}$$

where X_t, m_t, *and* h_t are the inputs at time t, cell update, and output at time t, respectively; W_i, W_f, W_C, W_o are the input gate weight matrices; and b_i, b_f, b_C, b_o are the bias matrices with respect to W_i, W_f, W_C, W_o matrices.

3) PROPOSED MODEL

The proposed model combines the DCNN–BiLSTM network and AM (CBiAM), as shown in Fig.7. 1DCNN–BiLSTM layers create the sequential data (h1, h2,…,hE), which are then fed into the AM [34]. In the attention layer, the AM will calculate the query matrix Q, key matrix K, and value matrix V based on the data input from the 1DCNN-BiLSTM layers. The query and key matrices are used to compute the attention weight of each hidden layer ($\alpha_1, \alpha_2, \ldots, \alpha_E$), which indicates the relevance of each element in the input sequence. The AM output is a weighted sum (Z) calculated by the attention weight and value matrix V, as shown in (11). Finally, the output of the AM is fed into a fully connected layer, followed by the softmax function. The architecture of the proposed model is shown in Table 2

$$Z\left(Q, K, V\right) = Softmax\left(\frac{QK^T}{\sqrt{D_M}}\right)V \tag{11}$$

where D_M is the dimensionality of the input sequence vector

TABLE 2. The architecture of the proposed model.

Number	Name of layers		Value
1	Convolution	Filter	16
		Kernel_size	5
		Padding	1
		Stride	Causal
2	Batch Normalization		
3	Leaky ReLU		0.1
4	Bidirectional_LSTM	Neuron	32
5	Attention Mechanism		
6	Dropout	Ratio	0.5
7	Dense	Neuron	5
	Activation	Softmax	

C. EXPERIMENTAL SETUP AND EVALUATION PARAMETERS

1) INPUT AND OUTPUT FEATURES

The CBiAM model uses a sequence of data to predict a cyclist's action; therefore, the input of model was shaped as sequential data. The number of sequences of raw dataset is defined in (12) and (13):

$$S_i = \left[\mathrm{Mea}_1^i, \mathrm{Med}_1^i, \ldots, \mathrm{St}_9^i\right]_{i=1,}^{N} \tag{12}$$

$$N = \frac{t - w}{w * o} + 1, \tag{13}$$

where S_i is the input sample *i-th*, N is the number of input samples, which is created from the total number of sample data points (t), w is the length of each sample, and o is the nonoverlapping ratio (Fig. 8). To maintain the detection time within 5s, the length of each sample w is chosen as 5 [8].

2) EVALUATION PARAMETERS

The performance of the CBiAM model was compared with other architectures based on four evaluation metrics: accuracy, precision, recall, and F1-score. Accuracy is the total number of correct predictions divided by the total number

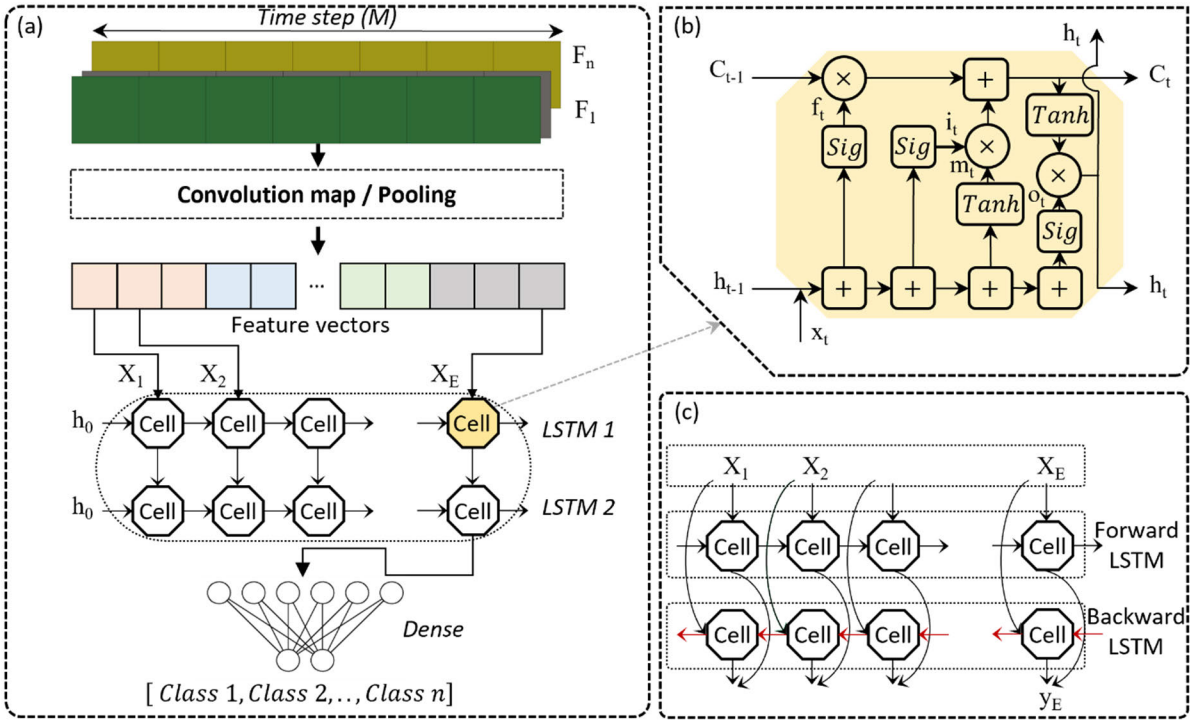

FIGURE 6. 1DCNN–BiLSTM model architecture.

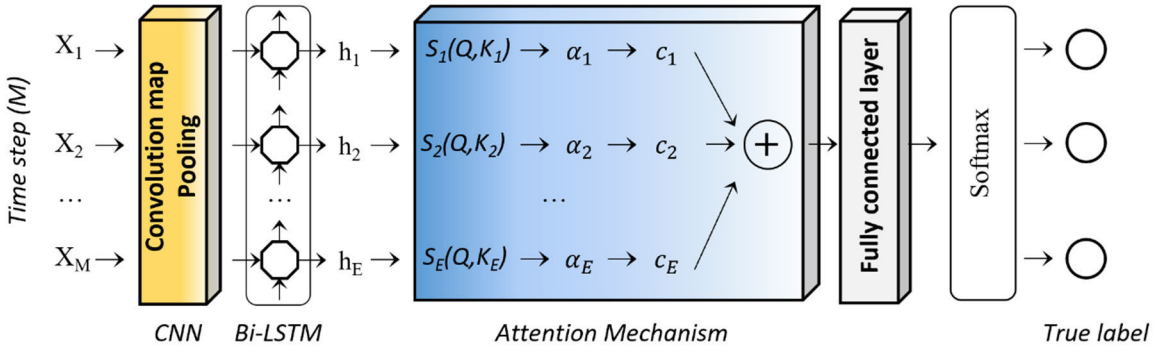

FIGURE 7. Architecture of the proposed CNN–BiLSTM model with attention mechanism.

of examples in the test set (14). Precision is the number of correct positive predictions divided by the total number of positive predictions (15). Recall is the ratio of the number of correct positive predictions to the sum of true positives and false negatives (16). Finally, the F1-score, which considers both precision and recall, is the idea parameter to compare the classification performance between models, as shown in (17) [21].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F_1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{17}$$

where TP, TN, FP, and FN refer to the true positive, true negative, false positive, and false negative, respectively.

3) HYPERPARAMETER SETTING

To validate the effectiveness of the CBiAM model, the proposed model was trained using different hyperparameters such as Lr, Bz, and nonoverlapping sliding windows. Five nonoverlapping ratios (o = 20%, 40%, 60%, 80%, and 100%) were used herein. The percentage indicates how far the sliding window travels relative to the first sample. In other words, the sliding windows will slide one step to the left to create a new sample if the first level of the sliding window is used (20%). Similarly, if the fifth level of the sliding window is used (100%), the next sample will contain the next five samples without using data from the previous sample, as shown in Fig. 8. Table 3 lists the hyperparameters used for training the proposed model.

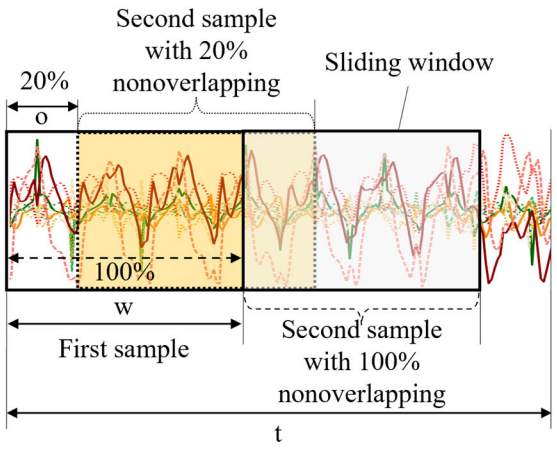

FIGURE 8. Sample construction for the input of the proposed model.

TABLE 3. Hyperparameters used to train the proposed model.

Hyperparameter	Description	Values
Lr	Learning rate	0.01, 0.001, 0.0001
Bz	Batch size	64, 128, 256
o	Level of nonoverlapping ratio	20%, 40% 60%, 80%, 100%
Ep	Number of Epochs	100
w	Sliding window	5
Op	Optimizer	Adam
LosFC	Loss function	Cross-entropy

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed model was trained using different hyperparameters to find the best model using the CySa dataset. Then, the proposed model was compared with various models that have shown good classification performance. Finally, the proposed model was trained on various public datasets and its performance was compared with other models to demonstrate its robustness.

A. IMPACT OF DIFFERENT PARAMETERS ON THE PROPOSED MODEL

As shown in Fig. 9(a), the training and validation losses decrease gradually, implying that the network is effectively learning without overfitting. The CBiAM model efficiently distinguishes the cyclist activities (Stan, Sit, Wal, Cyl, and Fal) from the CySa dataset (Table 4). The F1-scores and accuracies vary according to the Bz, Lr, and level of nonoverlapping ratio. Specifically, the CBiAM model yielded high F1-scores and accuracies when trained with an Lr of 0.01 and a Bz of 64, which is an almost 100% F1-score classification for five ratios. In other hyperparameters, the CBiAM model exhibited a lower prediction performance such as F1-scores of 0.9663 (Bz of 256, Lr of 0.01, and ratio of 80%) and 0.9166 (Bz of 256, Lr of 0.0001, and ratio of 100%). This error is possibly because the CySa dataset contains similar features of static activities (Stan and Sit) that do not have significant oscillations in the sensor readings. For dynamic activities (Wal, Cyl, and Fal), the CBiAM model yielded accurate predictions due to significant changes in accelerometer and gyroscope values across different directions. Thus, the CBiAM model recognizes these actions with high accuracy.

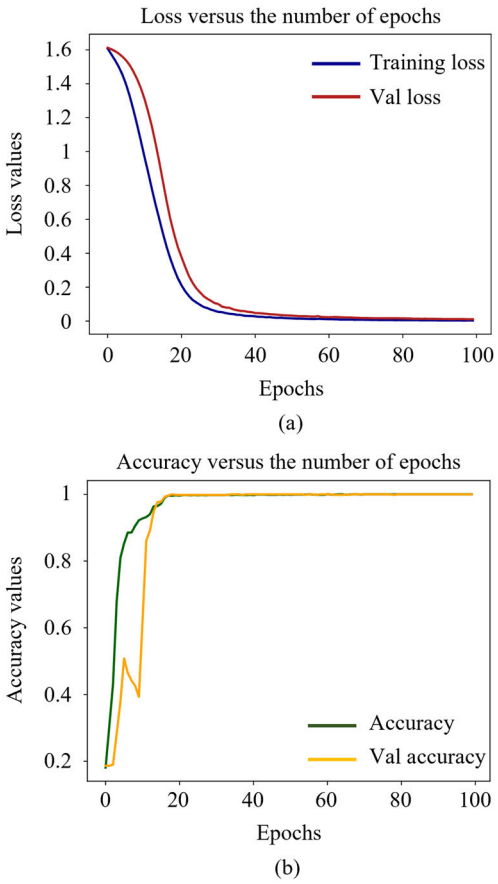

FIGURE 9. Training loss/accuracy and validation loss/accuracy of the proposed model trained on a Bz of 64 and Lr of 10^{-3}, (a) training and validation loss, (b) training and validation accuracy.

B. COMPARISION WITH OTHER CLASSIFICATION METHODS

The F1-score of the CBiAM model was compared with those of the existing architectures for HAR, such as KNN, SVM, ANN, 1DCNN, LSTM, and BiLSTM. All models were trained on the CySa dataset; Table 5 shows the best F1-scores for all models. Overall, all methods performed well on the CySa dataset. Particularly, DL methods, such as 1DCNN, LSTM, and BiLSTM, yielded better prediction results than the traditional learning methods such as KNN and SVM. Interestingly, the combination of LSTM/BiLSTM and 1DCNN models showed poor prediction performance, with F1-scores of 80.79% and 96.65%, respectively, compared to the single LSTM or single BiLSTM models. However, the CBiAM model outperformed the single LSTM and BiLSTM models as well as the classification models with an F1-score of 100%

TABLE 4. F1-score of The proposed model on the cysa dataset.

Ratio (%)	Batch size and Learning rate								
	0.01			0.001			0.0001		
	64	128	256	64	128	256	64	128	256
20%	0.994	1	1	0.9987	0.9981	0.9994	0.9968	0.9963	0.9939
40%	0.9975	0.9987	0.9987	0.9976	0.9938	0.9987	0.9962	0.9962	0.9914
60%	1	0.9963	0.9963	0.9981	0.9982	1	1	0.9944	0.9943
80%	1	1	0.9663	0.9975	0.9949	0.9975	0.9958	0.9949	0.9387
100%	1	1	1	1	1	0.9917	1	0.9937	0.9166

TABLE 5. Experimental results of the activity identification using different classification methods.

Architecture	F1-score
KNN [35]	98.39%
SVM [36]	97.93%
ANN [37]	99.27%
1DCNN [17]	99.39%
1DCNN [16]	99.71%
LSTM–1DCNN [21]	80.79%
BiLSTM [39]	99.09%
1DCNN–BiLSTM [38]	96.65%
CBiAM model	**100%**

TABLE 6. Public datasets.

Datasets	Samples	Features	Classes	Ratios
UCI-HAR	10,229	9	6	7:3
PAMAP2	1,942,872	52	12	8:2
Oppor_LLL	407,322	221	4	8:2
Oppor_HLA	375,901	221	5	8:2
MOTIONSENSE	1,412,865	12	6	8:2
WISDM	1,098,209	3	6	8:2

C. CLASSIFICATION PERPORMANCE RESULTS WITH PUBLIC DATASETS

The CBiAM model was trained using different public datasets to evaluate its classification performance; then, its performance was compared with those of the state-of-the-art models. Five public datasets, namely UCI-HAR, Opportunity, PAMAP2, WISDM, and MOTIONSENSE, were used. Table 6 shows the number of samples, features, classes, and the train/test ratio of each dataset.

1) PUBLIC DATASET DESCRIPTION

a: UCI-HAR

UCI-HAR [40] is a widely used public dataset. The smartphones were mounted on the waist of 30 participants who performed six activities, such as standing, sitting, lying down, walking, walking downstairs, and walking upstairs. The app installed on the smartphone recorded information from the body accelerometer and gyroscope.

b: OPPORTUNITY

The Opportunity dataset contains data on a set of specific activities, including cleaning the table, preparing a sandwich, opening and closing the dishwasher, and drinking coffee. It was created as part of a project aimed at developing methods for recognizing and understanding human activities using wearable sensors. To capture the participants' movements, 12 accelerometers were attached to different parts of their bodies. A total of 12 participants were involved in the experiments, and all the actions were labeled in approximately 57 h. Two types of opportunity annotations are widely used: (1) low-level locomotion (LLL) [41], which comprises four classes, namely stand, sit, walk, and lie, and high-level activities (HLA), which comprises five classes, namely relaxing, coffee time, early morning, clean up, and sandwich [42]. The LLL used herein was similar to that used in a previous study [41].

c: MOTIONSENSE

The MOTIONSENSE dataset was created by researchers at the MIT Media Lab, and it contained data on six activities performed by 24 participants, such as climbing upstairs, walking downstairs, sitting, standing, walking, and jogging, within the same environment and conditions [43]. The accelerometer and gyroscope sensors in smartwatches were used to measure attitude, acceleration, gravity, and rotation rate. Specifically, a total of 1,412,865 samples were collected, capturing 12 features: attitude-roll, attitude-pitch, attitude-yaw, gravity-x, gravity-y, gravity-z, rotation rate-x, rotation rate-y, rotation rate-z, acceleration-x, acceleration-y, and acceleration-z.

d: PAMAP2

The PAMAP2 dataset contained data on 18 physical activities performed by nine participants (eight male and one female), namely (1) lying, (2) sitting, and (3) standing. Three inertial measurement units (IMUs) were attached to the hand, chest, and ankle of the participants. The IMUs had a sampling frequency of 100 Hz, i.e., data were collected every 0.01 s. The total number of features in the dataset was 52. The recorded data were obtained from the accelerometer, gyroscope, magnetometer, temperature, electromyography, and heart rate measurements. Typically, six optional activities were removed from the final dataset, resulting in 12 classes being retained [44].

TABLE 7. Performance of the proposed model on public datasets.

Datasets	Proposed model	Other studies	Architectures
UCI-HAR	**97.51**	95.78 [21], 95.48 [38], 96.37 [46], 91.83 [51], 96.2 [56]	[21], [48] LSTM-CNN, [24] CNN-GRU
PAMAP2	**99.82**	94.29 [46], 93.21 [41], 96.7 [49], 91.25 [51], 94.16 [52], 96.35 [53], 94.76 [54], 90.03 [55], 95.27 [56], 96.67 [60]	[38], [46] 1DCNN–BiLSTM [41], [47], [49], [53], [57] 1DCNN [56], [60] CNN-GRU
Opportunity (LLL)	**94.72**	89.67 [41], 88.7 [47], 93 [48], 86.37 [55]	[58] LSTM [42] Probabilistic interval-based model
Opportunity (HLA)	**99.75**	98 [42], 96 [50]	[50] The maximum entropy learning [51] Random Forest classifier
WISDM	**97.67**	95.85 [21], 97.64 [24],96.05 [46], 93.32 [57], 94.21 [56], 94 [58], 96.5 [59], 96.41 [60]	[52] Kernel Sliding Perceptron [54] Bayesian optimization framework [55] Multilevel residual network with attention
MOTIONSENSE	**87.05**	79.86 [20], 85.15 [48]	[59] U-Net

[1]https://github.com/E2ELearning/Cycling_Monitoring_Applications

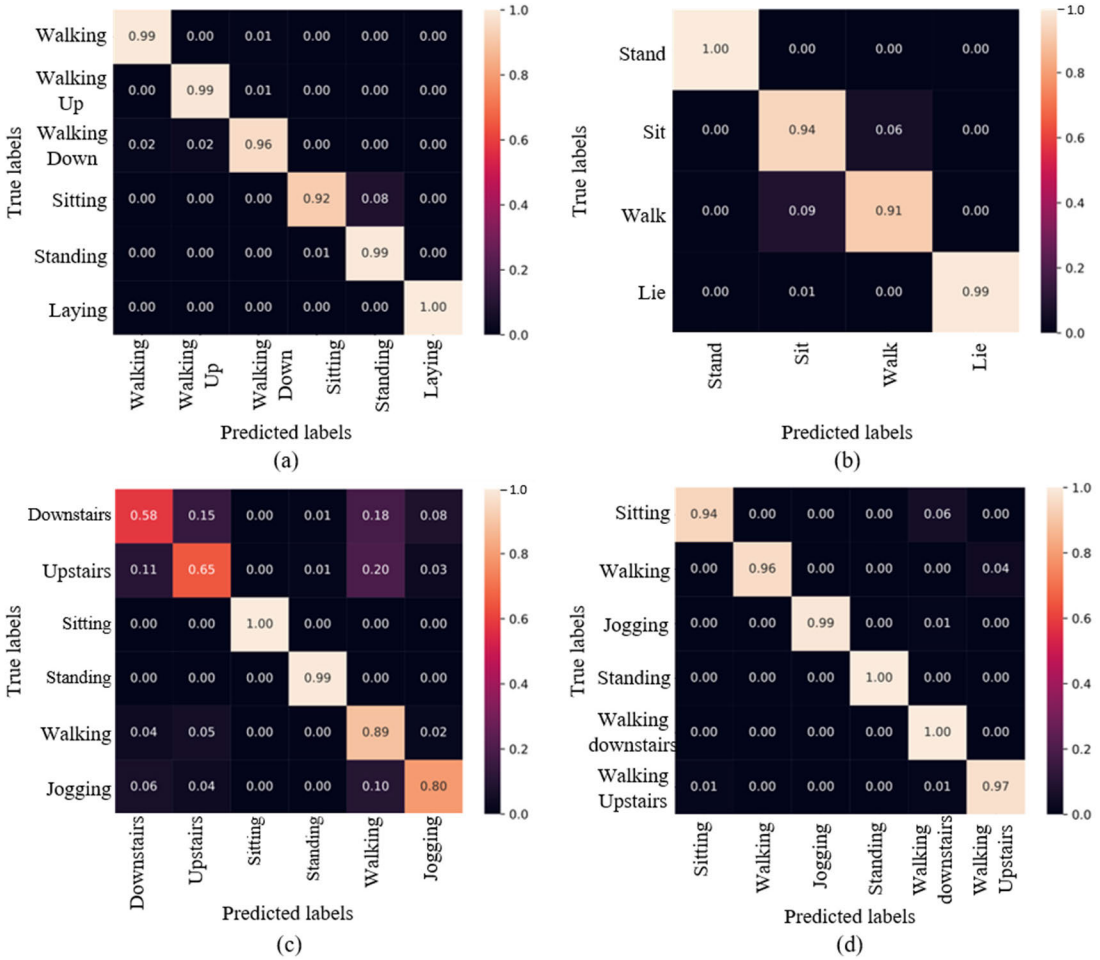

FIGURE 10. Confusion matrix of the proposed model on validation using the (a) UCI-HAR, (b) Opport_LLL, (c) MOTIONSENSE, and (d) WISDM datasets.

e: WISDM

The WISDM dataset contained data on six activities performed by 36 participants, (1) sitting, (2) walking, (3 jogging, (4) standing, (5) walking downstairs, and (6) climbing upstairs, and was created by Kwapisz et al. The participants used an Android phone and placed it in their front leg pockets to record the data at a sampling frequency of 20 Hz [45]. To maintain data quality, a dedicated individual supervised the data collection process. The most prevalent activity was ''walking,'' with 424,400 samples, followed by ''jogging''

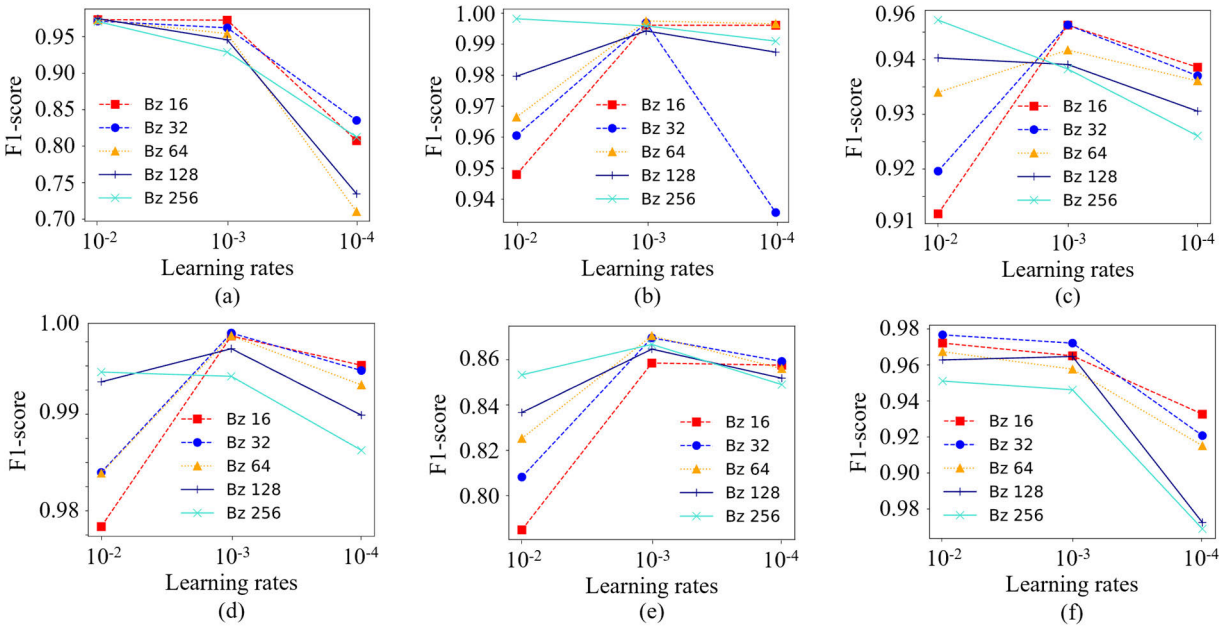

FIGURE 11. Training performance of the CBiAM model on different batch sizes and learning rates. (a) UCI-HAR, (b) PAMAP2, (c) Opport_LLL, (d) Opport_HLA, (e) MOTIONSENSE, and (f) WISDM datasets.

with 342,177 samples. "Standing" comprised the least number of samples (48,395), accounting for 4.4% of the dataset.

2) COMPARATIVE RESULTS

The proposed architecture was evaluated using five public datasets; the weighted F1-scores for the best predictions listed in Table 7. Compared with other architectures, such as LSTM [58], 1DCNN [41], [47], [49], [53], [57], 1DCNN–LSTM [21], [48], and 1DCNN–BiLSTM [38], [46], the CBiAM model was efficient in handling complex activities. Notably, the proposed network demonstrated superior classification results on the UCI-HAR dataset (97.52%) compared to LSTM [58], CNN-GRU [56], and 1DCNN–BiLSTM [46]. Similarly, the AM in the proposed model exhibited better prediction performance on the Opportunity (LLL) dataset compared to 1DCNN [41] and 1DCNN–LSTM [48], achieving an F1-score of 94.72%. Remarkable F1-scores of 99.82% and 97.67% were achieved on the PAMAP2 and WISDM datasets respectively, highlighting the superiority of the CBiAM model over the state-of-the-art approaches.

In the MOTIONSENSE dataset, the proposed model yielded a high classification accuracy, with a classification rate of 87.05%, surpassing the results of previous studies. However, due to the inherent variability in data collection (where 24 participants performed various actions, introducing substantial noise and variance), achieving high classification accuracy was challenging, as similarly observed in previous studies [20], [48]. For in-depth analysis, the confusion matrix of the MOTIONSENSE dataset was examined. The network exhibited lower accuracy in distinguishing between the actions "walking downstairs" and "climbing upstairs" because these actions share similar gestures and thus may appear similar, as shown in Fig. 10(c).

To achieve optimal classification for each dataset, the proposed model was trained with different Lr and Bz. The recognition performance varied based on each combination of the Bz and Lr, as shown in Fig. 11. For instance, both the UCI-HAR and WISDM datasets yielded favorable prediction outcomes when trained with a smaller Lr (10^{-2}), as shown in Figs. 11(a) and (f). In contrast, the PAMAP2, Opportunity, and MOTIONSENSE datasets achieved improved results with an Lr of 10^{-3}, as shown in Figs. 11(b)–(e).

Furthermore, we identified that Bz impacted the performance of the model. Smaller Bz yielded higher classification accuracy for the UCI-HAR and WISDM datasets, whereas larger Bz yielded higher classification accuracy for other datasets (Opportunity, PAMAP2, and MOTIONSENSE).

We have observed that while the proposed model achieved a perfect prediction on the proposed Cycling Safety (CySa) dataset and delivered promising results on most public datasets (PAMAP2, Opportunity, MOTIONSENSE, and WISDM) in comparison to previous studies, the combination of 1DCNN, BiLSTM, and the Attention mechanism can lead to an increase in the model's size, resulting in a larger memory space occupation in mobile applications. Nevertheless, with the advancements in mobile phone technology, modern devices can accommodate substantial amounts of data and support the installation of resource-intensive applications.

V. CONCLUSION AND FUTURE WORKS

In this study, the CySa dataset was presented, which is a comprehensive collection of cyclists' actions, such as walking

with a bicycle, sitting, cycling, and even falling. Additionally, a DL model, referred to as the 1DCNN–BiLSTM with an AM (CBiAM) model, was designed to accurately identify the actions of individual cyclists in cycling groups. The performance of the CBiAM model was verified via three experiments conducted using the CySa dataset. Therein, the impact of sliding window size, Bz, and Lr on the performance was evaluated. The performance of the proposed model was compared with those of the existing architectures trained on public datasets. The results indicate that the proposed model yields outstanding performance due to the integration of the AM and BiLSTM model, achieving perfect classification accuracy (100%) on the CySa dataset. Different architectures were trained on the same dataset, and the CBiAM model outperformed the traditional methods, such as SVM and AN, and advanced DL approaches, including 1DCNN, LSTM, and BiLSTM. Furthermore, the influence of sliding window size, Bz, and Lr on the model's performance was evaluated using the proposed dataset. The robustness of the CBiAM model was evaluated based on its enhanced recognition capabilities across various publicly available datasets, namely UCI-HAR, PAMAP2, Opportunity, MOTIONSENSE, and WISDM.

The proposed model exhibits favorable recognition performance on the CySa dataset; however, the memory consumption and prediction time across all architectures must be analyzed in future work. This consideration will pave the way for an optimal solution pertaining to memory utilization for the trained model, particularly in real-world applications on mobile devices. Furthermore, a similar ratio between the training and test sets of public datasets during the training phase will be used to ensure unbiased comparisons. To enhance the reliability of results, a fivefold cross-validation approach will also be considered.

REFERENCES

[1] L.-A. Leyland, B. Spencer, N. Beale, T. Jones, and C. M. van Reekum, "The effect of cycling on cognitive function and well-being in older adults," *PLoS ONE*, vol. 14, no. 2, Feb. 2019, Art. no. e0211779.

[2] N. C. McDonald, R. L. Steiner, W. M. Palmer, A. N. Bullock, V. P. Sisiopiku, and B. F. Lytle, "Costs of school transportation: Quantifying the fiscal impacts of encouraging walking and bicycling for school travel," *Transportation*, vol. 43, no. 1, pp. 159–175, Jan. 2016.

[3] C. Juhra, "Bicycle accidents-do we only see the tip of the iceberg: A prospective multi-centre study in a large German city combining medical and police data," *Injury*, vol. 43, no. 12, pp. 2026–2034, Dec. 2012.

[4] R. Utriainen, M. Pöllänen, and H. Liimatainen, "Road safety comparisons with international data on seriously injured," *Transp. Policy*, vol. 66, pp. 138–145, Aug. 2018.

[5] K. Aparna, H. D. Dayajanaki, D. Devu, S. P. Rajeev, and S. D. B. Sreeja, "Wearable sensors in daily life: A review," in *Proc. 9th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Mar. 2023, pp. 863–868.

[6] M. Elgendi, "Mobile and wearable systems for health monitoring," *Frontiers Digit. Health.*, vol. 5, Apr. 2023, Art. no. 1196103.

[7] M. Berchtold, M. Budde, D. Gordon, H. R. Schmidtke, and M. Beigl, "ActiServ: Activity recognition service for mobile phones," in *Proc. Int. Symp. Wearable Comput. (ISWC)*, Oct. 2010, pp. 1–8.

[8] M. Straczkiewicz, P. James, and J.-P. Onnela, "A systematic review of smartphone-based human activity recognition methods for health research," *NPJ Digit. Med.*, vol. 4, no. 1, p. 148, Oct. 2021.

[9] G. S. Carvalho and F. O. Silva, "Performance analysis of relative GPS positioning as function of communication latency," in *Proc. LARS-SBR-WRE*, Oct. 2021, pp. 252–257.

[10] A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell, "Activity recognition using a single accelerometer placed at the wrist or ankle," *Med. Sci. Sports Exercise*, vol. 45, no. 11, pp. 2193–2203, 2013.

[11] A. Anjum and M. U. Ilyas, "Activity recognition using smartphone sensors," in *Proc. IEEE 10th Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, Jan. 2013, pp. 914–919.

[12] S. Mehrang, "Human activity recognition using a single optical heart rate monitoring wristband equipped with triaxial accelerometer," in *Proc. EMBEC NBC*, Cham, Switzerland: Springer, 2017, pp. 587–590.

[13] P. Li, Y. Wang, Y. Tian, T.-S. Zhou, and J.-S. Li, "An automatic user-adapted physical activity classification method using smartphones," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 3, pp. 706–714, Mar. 2017.

[14] M.-C. Kwon and S. Choi, "Recognition of daily human activity using an artificial neural network and smartwatch," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–9, Jun. 2018.

[15] G. Li, L. Huang, and H. Xu, "IWalk: Let your smartphone remember you," in *Proc. 4th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Jul. 2017, pp. 414–418.

[16] B. Zhou, J. Yang, and Q. Li, "Smartphone-based activity recognition for indoor localization using a convolutional neural network," *Sensors*, vol. 19, no. 3, p. 621, Feb. 2019.

[17] R. Zhu, Z. Xiao, Y. Li, M. Yang, Y. Tan, L. Zhou, S. Lin, and H. Wen, "Efficient human activity recognition solving the confusing activities via deep ensemble learning," *IEEE Access*, vol. 7, pp. 75490–75499, 2019.

[18] A. W. Sardar, F. Ullah, J. Bacha, J. Khan, F. Ali, and S. Lee, "Mobile sensors based platform of human physical activities recognition for COVID-19 spread minimization," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105662.

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

[20] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "Physical activity recognition based on deep learning using photoplethysmography and wearable inertial sensors," *Electronics*, vol. 12, no. 3, p. 693, Jan. 2023.

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

[22] F. M. Noori, M. Riegler, M. Z. Uddin, and J. Torresen, "Human activity recognition from multiple sensors data using multi-fusion representations and CNNs," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 2, pp. 1–19, May 2020.

[23] Y. A. Andrade-Ambriz, S. Ledesma, M.-A. Ibarra-Manzano, M. I. Oros-Flores, and D.-L. Almanza-Ojeda, "Human activity recognition using temporal convolutional neural network architecture," *Expert Syst. Appl.*, vol. 191, Apr. 2022, Art. no. 116287.

[24] N. Dua, S. N. Singh, V. B. Semwal, and S. K. Challa, "Inception inspired CNN-GRU hybrid network for human activity recognition," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 5369–5403, Feb. 2023.

[25] Z. N. Khan and J. Ahmad, "Attention induced multi-head convolutional neural network for human activity recognition," *Appl. Soft Comput.*, vol. 110, Oct. 2021, Art. no. 107671.

[26] S. Mekruksavanich, A. Jitpattanakul, K. Sitthithakerngkiet, P. Youplao, and P. Yupapin, "ResNet-SE: Channel attention-based deep residual network for complex activity recognition using Wrist–Worn wearable sensors," *IEEE Access*, vol. 10, pp. 51142–51154, 2022.

[27] C. Han, L. Zhang, Y. Tang, S. Xu, F. Min, H. Wu, and A. Song, "Understanding and improving channel attention for human activity recognition by temporal-aware and modality-aware embedding," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[28] X. Yin, Z. Liu, D. Liu, and X. Ren, "A novel CNN-based bi-LSTM parallel model with attention mechanism for human activity recognition with noisy data," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, May 2022.

[29] J. Wannenburg and R. Malekian, "Physical activity recognition from smartphone accelerometer data for user context awareness sensing," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 47, no. 12, pp. 3142–3149, Dec. 2017.

[30] I. M. Pires, G. Marques, N. M. Garcia, F. Flórez-Revuelta, M. C. Teixeira, E. Zdravevski, S. Spinsante, and M. Coimbra, "Pattern recognition techniques for the identification of activities of daily living using a mobile device accelerometer," *Electronics*, vol. 9, no. 3, p. 509, Mar. 2020.

[31] E. G. Zimbelman and R. F. Keefe, "Development and validation of smartwatch-based activity recognition models for rigging crew workers on cable logging operations," *PLoS ONE*, vol. 16, no. 5, May 2021, Art. no. e0250624.

[32] M. Omar, S. Choi, D. Nyang, and D. Mohaisen, "Robust natural language processing: Recent advances, challenges, and future directions," *IEEE Access*, vol. 10, pp. 86038–86056, 2022.
[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[34] A. Vaswani, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[35] M. Arif, M. Bilal, A. Kattan, and S. I. Ahamed, "Better physical activity classification using smartphone acceleration sensor," *J. Med. Syst.*, vol. 38, no. 9, p. 95, Sep. 2014.
[36] J. Rabbi, M. T. H. Fuad, and M. A. Awal, "Human activity analysis and recognition from smartphones using machine learning techniques," 2021, *arXiv:2103.16490*.
[37] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, D. Makris, and G. Farias, "Fall detection and activity recognition using human skeleton features," *IEEE Access*, vol. 9, pp. 33532–33542, 2021.
[38] Y. J. Luwe, C. P. Lee, and K. M. Lim, "Wearable sensor-based human activity recognition with hybrid deep learning model," *Informatics*, vol. 9, no. 3, p. 56, Jul. 2022.
[39] S. Yu and L. Qin, "Human activity recognition with smartphone inertial sensors using bidir-LSTM networks," in *Proc. 3rd Int. Conf. Mech., Control Comput. Eng. (ICMCCE)*, Huhhot, China, Sep. 2018, pp. 219–224.
[40] 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. ESANN*, 2013, pp. 1–6.
[41] F. M. Rueda, R. Grzeszick, G. Fink, S. Feldhorst, and M. ten Hompel, "Convolutional neural networks for human activity recognition using body-Worn sensors," *Informatics*, vol. 5, no. 2, p. 26, May 2018.
[42] L. Liu, L. Cheng, Y. Liu, Y. Jia, and D. S. Rosenblum, "Recognizing complex activities by a probabilistic interval-based model," in *Proc. AAAI*, 2016, pp. 1266–1272.
[43] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Protecting sensory data against sensitive inferences," in *Proc. 1st Workshop Privacy Design Distrib. Syst.*, New York, NY, USA, Apr. 2018, pp. 1–6.
[44] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Proc. 16th Int. Symp. Wearable Comput.*, Jun. 2012, pp. 108–109.
[45] 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.
[46] 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, Aug. 2021.
[47] R. Yao, G. Lin, Q. Shi, and D. C. Ranasinghe, "Efficient dense labelling of human activity sequences from wearables using fully convolutional networks," *Pattern Recognit.*, vol. 78, pp. 252–266, Jun. 2018.
[48] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.
[49] M. Gil-Martín, R. San-Segundo, F. Fernández-Martínez, and R. de Córdoba, "Human activity recognition adapted to the type of movement," *Comput. Electr. Eng.*, vol. 88, Dec. 2020, Art. no. 106822.
[50] Z. Huo, A. PakBin, X. Chen, N. Hurley, Y. Yuan, X. Qian, Z. Wang, S. Huang, and B. Mortazavi, "Uncertainty quantification for deep context-aware mobile activity recognition and unknown context discovery," 2020, *arXiv:2003.01753*.
[51] S. B. U. D. Tahir, A. B. Dogar, R. Fatima, A. Yasin, M. Shafiq, J. A. Khan, M. Assam, A. Mohamed, and E.-A. Attia, "Stochastic recognition of human physical activities via augmented feature descriptors and random forest model," *Sensors*, vol. 22, no. 17, p. 6632, Sep. 2022.
[52] M. Gochoo, S. B. U. D. Tahir, A. Jalal, and K. Kim, "Monitoring real-time personal locomotion behaviors over smart indoor-outdoor environments via body-Worn sensors," *IEEE Access*, vol. 9, pp. 70556–70570, 2021.
[53] M. Gil-Martín, R. San-Segundo, F. Fernández-Martínez, and J. Ferreiros-López, "Time analysis in human activity recognition," *Neural Process. Lett.*, vol. 53, no. 6, pp. 4507–4525, 2021.
[54] M. A. Awal, M. K. Hasan, M. A. Rahman, and M. A. Alahe, "Optimization of daily physical activity recognition with feature selection," in *Proc. 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2019, pp. 1–6.
[55] M. A. A. Al-qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 144–152, Jan. 2023.
[56] 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, Mar. 2021.
[57] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Appl. Soft Comput.*, vol. 62, pp. 915–922, Jan. 2018.
[58] S. W. Pienaar and R. Malekian, "Human activity recognition using LSTM-RNN deep neural network architecture," in *Proc. IEEE 2nd Wireless Afr. Conf. (WAC)*, Aug. 2019, pp. 1–5.
[59] Y. Zhang, Z. Zhang, Y. Zhang, J. Bao, Y. Zhang, and H. Deng, "Human activity recognition based on motion sensor using U-Net," *IEEE Access*, vol. 7, pp. 75213–75226, 2019.
[60] L. Lu, C. Zhang, K. Cao, T. Deng, and Q. Yang, "A multichannel CNN-GRU model for human activity recognition," *IEEE Access*, vol. 10, pp. 66797–66810, 2022.

VAN SY NGUYEN received the B.S. degree from Vietnam National University Ho Chi Minh City (VNU-HCM)-Ho Chi Minh City University of Technology, in 2017, and the integrated M.S. and Ph.D. degrees from Korea Aerospace University, in 2023. He was a Postdoctoral Research Associate with the Department of Convergence and Fusion System Engineering, Kyungpook National University, South Korea. He is currently a Postdoctoral Research Associate with the Department of Mechanical and Aerospace Engineering, University of Central Florida, USA. His research interests include robotics, mechanisms and control, bio/medical application robots, computer vision, and deep learning.

HYUNSEOK KIM received the B.S. degree in electronics engineering from Dong-A University, Busan, Republic of Korea, in 2001, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2004 and 2014, respectively. From 2001 to 2003, he was with Samsung Electronics Company Ltd. From 2004 to 2009, he was with LG Electronics Inc. From 2011 to 2022, he was with the Electronics and Telecommunications Research Institute (ETRI). Since 2022, he has been an Assistant Professor with the Department of Computer Engineering, Dong-A University. His main research interests include reinforcement learning, swarm intelligence, and developing innovative behavior intelligence in multi-agent systems.

DONGJUN SUH (Member, IEEE) received the Ph.D. degree from the Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology (KAIST), in 2014. From 2015 to 2018, he was a Senior Researcher with the Korea Institute of Science and Technology Information (KISTI). From 2014 to 2015, he was a Research Assistant Professor with the KAIST Institute for Information Technology Convergence. He is currently an Associate Professor with the Department of Convergence and Fusion System Engineering, Kyungpook National University (KNU), South Korea. His research interests include big data and smart control systems using predictive analytics based on machine learning and deep learning.