informatics

MDPI

Article

Wearable Sensor-Based Human Activity Recognition with Hybrid Deep Learning Model

Yee Jia Luwe, Chin Poo Lee * and Kian Ming Lim

Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia; 1181101480@student.mmu.edu.my (Y.J.L.); kmlim@mmu.edu.my (K.M.L.)
* Correspondence: cplee@mmu.edu.my

Abstract: It is undeniable that mobile devices have become an inseparable part of human's daily routines due to the persistent growth of high-quality sensor devices, powerful computational resources and massive storage capacity nowadays. Similarly, the fast development of Internet of Things technology has motivated people into the research and wide applications of sensors, such as the human activity recognition system. This results in substantial existing works that have utilized wearable sensors to identify human activities with a variety of techniques. In this paper, a hybrid deep learning model that amalgamates a one-dimensional Convolutional Neural Network with a bidirectional long short-term memory (1D-CNN-BiLSTM) model is proposed for wearable sensor-based human activity recognition. The one-dimensional Convolutional Neural Network transforms the prominent information in the sensor time series data into high level representative features. Thereafter, the bidirectional long short-term memory encodes the long-range dependencies in the features by gating mechanisms. The performance evaluation reveals that the proposed 1D-CNN-BiLSTM outshines the existing methods with a recognition rate of 95.48% on the UCI-HAR dataset, 94.17% on the Motion Sense dataset and 100% on the Single Accelerometer dataset.

Keywords: human activity recognition; convolutional neural network; long short-term memory; wearable sensor

Citation: Luwe, Y.J.; Lee, C.P.; Lim, K.M. Wearable Sensor-Based Human Activity Recognition with Hybrid Deep Learning Model. *Informatics* **2022**, *9*, 56. https://doi.org/10.3390/informatics9030056

Academic Editor: Guangjie Han

Received: 16 June 2022
Accepted: 18 July 2022
Published: 31 July 2022

1. Introduction

In recent times, the rapid development of human activity recognition is revealing massive real-world implementations in human's daily lives. For instance, Active and Assisted Living systems for smart homes, healthcare and eldercare monitoring systems and Virtual Reality or Tele-Immersion applications [1]. Due to rapid evolution and enthusiasm, this area has attracted a substantial number of studies on various approaches.

In general, human activity recognition systems follow a standard sequence of tasks. The very early task is to select a suitable tool to monitor and record the individual's movements. Depending on the selected tool, the kind of information to be collected and processed and thereby the feature extraction approach are determined. After the feature extraction approach is decided, the final task is to develop a suitable classifier to infer the activity class from the extracted features. There are often two kinds of data collection tools in human activity recognition systems, which are video-based [2] and sensor-based.

In this paper, a hybrid one-dimensional Convolutional Neural Network with a bidirectional long short-term memory (1D-CNN-BiLSTM) model is devised for sensor-based human activity recognition. The 1D-CNN model is leveraged to learn the salient features from the sensor data associated with each activity class. Subsequently, the BiLSTM model encodes the long-range dependencies of the features by the gating mechanisms.Hyperparameter tuning is performed with a grid search to determine the optimal settings of the model. The main contributions of this paper are:

- A hybrid 1D-CNN-BiLSTM model that amalgamates the strengths of both a CNN and a BiLSTM is proposed for human activity recognition.

- The 1D-CNN discovers the high-level discriminative features that represent the activity-specific characteristics, thus suppressing the impacts of the outliers and insignificant sensor data.
- The BiLSTM encodes the bidirectional long-range dependencies in the features by the gating mechanisms. The BiLSTM is efficient in alleviating the information loss and vanishing gradient problems.
- The hyperparameter tuning on an optimizer, the BiLSTM merge mode and the batch size are conducted to empirically determine the optimal values for the hyperparameters.
- The comparative study of the performance of the proposed 1D-CNN-BiLSTM with the existing methods using three human activity recognition datasets, namely an UCI-HAR dataset, a Motion Sense dataset and a Single Accelerometer dataset is carried out.

The rest of the paper is organized as follows. Section 2 provides a review of the existing works in the human activity recognition field. A detailed description of the proposed 1D-CNN-BiLSTM model is presented in Section 3. The datasets used in the experiments are described in Section 4. Section 5 provides the hyperparameter tuning results for the optimal model settings. The experimental results and confusion matrices are analyzed in Section 6. Finally, the conclusions are drawn in Section 7.

2. Related Works

This section reviews some existing deep learning models for human activity recognition [3–9].

In early work, Murad and Pyun (2017) [10] proposed Long Short-Term Memory (LSTM), BiLSTM and cascaded LSTM for human activity recognition. The LSTM and BiLSTM model consisted of an input layer, LSTM/BiLSTM layers and an output layer. The cascaded LSTM model comprised an input layer and a BiLSTM layer, followed by LSTM layers and an output layer. The LSTM model achieved accuracies of 97.8% on the USC-HAD dataset. The BiLSTM model recorded 92.5% accuracy on the Opportunity dataset, whereas the cascaded LSTM model obtained 94.1% accuracy on the Daphnet FOG dataset and 92.6% on the Skoda dataset.

Ignatov (2018) [11] conducted a comparison of a proposed shallow Convolutional Neural Network (CNN) framework with five layers to the existing solutions using a WISDM dataset and an UCI-HAR dataset. The obtained results showed that their proposed CNN model outstripped other CNN-based methods over the UCI-HAR dataset with an accuracy of 94.35%.

Moya et al. (2018) [12] introduced a CNN-IMU model for human activity recognition based on the inertial measurement unit (IMU). The CNN model was implemented in parallel branches with temporal convolutions to process and merge input sequences from IMUs individually. This is followed by a max pooling layer, a fully-connected layer and a classification layer. The CNN-IMU model achieved 92.24% accuracy on the Opportunity-Gestures dataset, 88.67% on the Opportunity-Locomotion dataset and 91.22% on the Pamap2 dataset.

In other work, Ferrari et al. (2019) [13] introduced the idea of comparing the k-Nearest Neighbour (k-NN) and Support Vector Machine (SVM) classifiers with handcrafted features to a seven-layered Residual Network (ResNet) for human activity recognition. After experimenting, the results showed that the ResNet deep learning model surpassed the performances of k-NN and SVM by obtaining an average accuracy of 92.94% across the datasets.

In a later work, Ragab et al. (2020) [14] developed a one-dimensional deep Convolutional Neural Network using Random Search (RS-1D-CNN). The central objective is to feed the input signals from the selected dataset into the proposed model and then carry out random search to come up with the most appropriate network connections and hyperparameter settings for model enhancement. The experiments on the UCI-HAR dataset showed that the RS-1D-CNN with an established 95.4% accuracy outperformed the other deep learning approaches.

In Zhao et al. (2018) [15], a Residual Bidirectional Long Short-Term Memory (Res-Bidir-LSTM) was introduced as the deep learning network framework for human activity recognition. The Res-Bidir-LSTM model showed an advancement in speed and efficacy of the temporal and spatial dimensions through the bidirectional integration of the forward states and backward states as well as the connections between the residual layer. The stacked cells also served as a time-saving approach for the prevention of the gradient-vanishing problem. The experimental results showed that the model attained a 93.60% accuracy and a 93.50% F1-score on the utilized UCI-HAR dataset and surpassed other deep learning methods such as Baseline LSTM [15], Bidir-LSTM [15] and Res-LSTM [15].

Another CNN-LSTM model was discussed in Mutegeki and Han (2020) [16]. The CNN-LSTM model comprised an input layer, four convolutional layers, a LSTM layer, a fully connected layer and a classification layer. The CNN-LSTM model recorded an accuracy of 92.13% on the UCI dataset and 99.06% on the iSPL dataset.

Ni et al. (2020) [17] utilized stacked denoising autoencoders (SDAE) to extract the features from the sensor data. There were two denoising autoencoders in the model with each autoencoder consisting of an input layer, a hidden layer and an output layer. The SDAE model obtained 97.15% accuracy on the smartphone dataset, 89.99% on the single accelerometer dataset and 95.26% on the UCI dataset.

An integration 1D-CNN and LSTM for human activity recognition was presented in Goh et al. (2021) [18]. The 1D-CNN was leveraged to learn high-level features from the sensor data, while the LSTM was used to encode the temporal dependencies of the features. The proposed model recorded an F1-score of 91.04% and 76.42% on the UCI-HAR and USC-HAD datasets, respectively.

Erdaş and Güney (2021) [19] proposed three models for human activity recognition, namely CNN, Convolutional LSTM (ConvLSTM) and 3D-CNN fed by ConvLSTM. The CNN model was built of an input layer, four convolutional layers, a dense layer and a classification layer. The ConvLSTM model had a similar architecture where there were an input layer, four convolutional LSTM layers, a dense layer and a classification layer. As for the 3D-CNN feb by ConvLSTM model, there was an input layer, four convolutional LSTM layers, a 3D-CNN layer, a dense layer and a classification layer. On the single accelerometer dataset, the CNN, ConvLSTM and 3D-CNN fed by ConvLSTM models yielded accuracies of 91.77%, 92.29% and 93.69%, respectively.

3. One-Dimensional Convolutional Neural Network with Bidirectional Long Short-Term Memory

This paper proposes a hybrid deep learning model, referred to as the "one-dimensional Convolutional Neural Network with Bidirectional Long Short-Term Memory (1D-CNN-BiLSTM)". It is known that CNNs are always the preferable models for diverse computer vision challenges. This is because CNNs can be constructed with numerous hidden layers and adjusted with many hyperparameter settings. These properties enable CNNs to learn the internal representation of different dimensions of signals for feature learning, such as images and videos. The uniform process can also be exploited on 1D signal data, in such cases as time series. Nevertheless, although the performance of 1D-CNN is already promising in feature learning, the versatility of 1D-CNN can be further explored with the incorporation of the other neural networks, intending to realize innovation with respect to speed and competence; this is where the Long Short-Term Memory (LSTM) comes in handy.

Figure 1 displays the architecture of the proposed 1D-CNN-BiLSTM model. The details of the 1D-CNN-BiLSTM are presented in Table 1, including the layer name as well as the hyperparameter settings of each layer. There are a total of 12 layers in the proposed 1D-CNN-BiLSTM model.

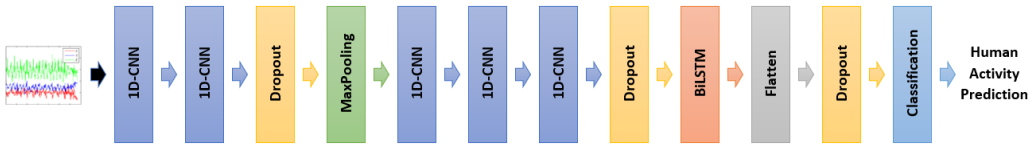

Figure 1. The architecture of the proposed 1D-CNN-BiLSTM model.

Table 1. Detailed architecture of 1D-CNN-BiLSTM.

Layer Name	Hyperparameter Settings
Conv1D	Kernel Size = 5, Strides = 1, Padding = Same, Activation = ReLU
Conv1D	Kernel Size = 5, Strides = 1, Padding = Same, Activation = ReLU
Dropout	Dropout Rate = 0.2
MaxPooling1D	Pool Size = 8, Strides = 1, Padding = Same
Conv1D	Kernel Size = 5, Strides = 1, Padding = Same, Activation = ReLU
Conv1D	Kernel Size = 5, Strides = 1, Padding = Same, Activation = ReLU
Conv1D	Kernel Size = 5, Strides = 1, Padding = Same, Activation = ReLU
Dropout	Dropout Rate = 0.2
BiLSTM	Activation = Tanh, Recurrent Activation = Sigmoid, Return Sequences = True, Recurrent Dropout = 0.0
Flatten	-
Dropout	Dropout Rate = 0.2
Classification	Activation = Softmax

3.1. 1-Dimensional Convolutional Neural Network

In this work, a 1D-CNN is leveraged to retrieve the features out of the time series sensor data and map the internal characteristics to various activity types. Unlike the classical machine learning methods that require manually handcrafting the features during the feature engineering process, the 1D-CNN is beneficial as the features are learned straight from the input data without any aid of manual feature engineering.

The first convolutional layer reads the multivariate time series with a specified length and width. The length is the value of time steps, while the width is the number of variables in the time series. The kernel will have the same width but different length as the time series. The first convolutional layer performs the convolution operations by multiplying the input time series with the filter matrix to obtain the high-level features. After that, the product from the multiplication will be totaled up and fed into a Rectified Linear Unit (ReLU) activation function. ReLU provides better gradient propagation and is less susceptible to vanishing gradient issues.

The next 1D convolutional layer will eventually convolve in a single channel. Under the convolutional process, there are three important hyperparameters to take note: kernel size, padding and stride. The kernel size in the proposed 1D-CNN-BiLSTM model is set to five which indicates the length of the sequential window. The kernel can only shift in one dimension along the axis of time steps. In addition, the dimensionality of the convolved feature can be lessened, increased or unchanged by specifying the type of padding to be used. In the proposed model, the same padding is fixed for every layer to keep the original dimension of the input time series. On the other hand, stride is the number of steps specified to move the filter along the time series. The stride is set as one in the proposed 1D-CNN-BiLSTM model.

3.2. Dropout Layer

Between the convolutional layers, three dropout layers with a dropout rate r of 0.2 are inserted to avoid overfitting. The dropout layers work by arbitrarily adjusting the input units to 0 at the dropout rate. In contrast, the input units that do not fit to 0 are scaled up using the expression of $\frac{1}{(1-r)}$ so that the total of the entire input units is unaffected.

3.3. 1D Max-Pooling Layer

The output from the convolutional layer will be fed into the succeeding layer, which is the 1D max-pooling layer. The 1D max-pooling layer is responsible to compute a maximum value for each time series vector against a spatial window with a regulated pool size. The spatial window is moved by the stride size. In the proposed 1D-CNN-BiLSTM model, the pool size is set as eight and the stride size is one. Since the same padding is used in the system, the output shape is acquired using the formula as follows:

$$\text{output shape} = \frac{\text{input shape}}{\text{stride}} \tag{1}$$

3.4. Bidirectional Long Short-Term Memory

The feature vector from the convolutional layer is passed into the bidirectional long short-term memory (Bi-LSTM). LSTM works out the common vanishing gradient problems in Recurrent Neural Networks by learning long-term sequences of actions or movements in human activities and the impact of initial dependencies in the related sequence [20]. In other words, LSTM not only reads the input time series at the current time step, but also captures the information from the time series that it perceived earlier using its hidden state.

However, unidirectional LSTM only handles time series based upon the preserved information in one direction, either the future or the past. In the real world, human activities are ceaseless, and the time series are being generated constantly. To handle long sequences of time series and prevent information loss, BiLSTM is a more appropriate option to be implemented. BiLSTM differs from unidirectional LSTM by training two LSTM layers instead of one, using the time series flowing in both directions, involving backwards and forwards, implying from past to future or vice versa. Therefore, both future and past information could be preserved which offers a supplementary context to the proposed network. This may lead to a faster and richer learning of the human activity recognition task.

3.5. Classification Layer

The feature vector from the Bi-LSTM layers is then flattened and passed into the classification layer. The classification layer is a fully connected layer that returns the prediction of all activity classes. In the classification layer, a Softmax activation function is applied to measure the probability distribution of the activity classes. The formula of the Softmax activation function (σ) is denoted as below:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2}$$

where $\vec{z}$ is the input vector, z_i are the components of the input vector where they can be any real values, and $\sum_{j=1}^{K} e^{z_j}$ is the normalization expression that squashes all outputs to the scale from 0 to 1, hence establishing an applicable probability distribution.

4. Datasets

Three publicly available datasets are used in the performance evaluation of the proposed 1D-CNN-BiLSTM for human activity recognition, namely the the UCI-HAR dataset [21], the Motion Sense dataset [22] and the Single Accelerometer dataset [23].

4.1. UCI-HAR Dataset

The UCI-HAR dataset contains six classes of activities from 30 subjects, including "walking", "walking up stairs", "walking down stairs", "sitting", "standing" and "laying". The time series of acceleration and angular velocity in the x, y and z axes were gathered through the accelerometer and gyroscope sensors engaged on the waist of the volunteers, respectively.

Several data pre-processing steps were performed on the raw sensor data, such as noise filtering and time series sampling with sliding windows of 2.56 s. The gravitational

and body motion parts from the acceleration time series were segregated from each other through the deployment of a Butterworth low-pass filter, and a 0.3 Hz isolation frequency was executed as well.

Additionally, feature vectors were extracted by computing mean and standard deviation for each window from the time and frequency domain. Lastly, the dataset was split into a training set and a testing set with the ratio of 7:3, resulting in 10,299 samples.

4.2. Motion Sense Dataset

The Motion Sense dataset includes six classes of activities from 24 subjects, including "upstairs", "downstairs", "sitting", "standing", "walking" and "jogging".

The time series of attitude, acceleration, gravity and rotation rate were collected using accelerometer and gyroscope sensors for 15 experiments carried out by each of the participants. There are 12 features available for each time series, involving roll, pitch and yaw for attitude, x, y and z for gravity, rotation rate and acceleration, respectively.

In the data pre-processing phase, all features and activity classes were included in the experiments to evaluate the performance of the proposed models. Finally, the dataset contains 1,412,865 samples in total. The dataset is randomly split into a training set and a testing set with the ratio of 8:2.

4.3. Single Accelerometer Dataset

The Single Accelerometer dataset was initially made up of eight classes of activities from 15 subjects, including "none", "working at computer", "standing up, walking and going up or down stairs", "standing", "walking", "going up or down stairs", "walking and talking with someone", "talking while standing". The time series of linear acceleration in the x, y and z axes were collected through the wearable accelerometer embedded on the participants' chests.

It is noticeable that the dataset contains samples without any activity (class "none"), or multiple activities (class "standing up, walking and going up or down stairs" and class "walking and talking with someone"). These three classes were eliminated from the dataset. Hence, the Single Accelerometer dataset now has only five classes of activities, namely "going up and down stairs", "working at computer", "talking while standing", "standing", and "walking". In addition, the missing values that existed within the dataset are also being removed. After data cleaning, there are 1,801,306 samples and the dataset is arbitrarily partitioned with the ratio of 8:2 for the the training set and the testing set.

The description of each dataset is summarized in Table 2.

Table 2. Summary of three datasets.

Dataset	Samples	Features	Classes	Ratio (Train:Test)
UCI-HAR	10,299	6	6	7:3
Motion Sense	1,412,865	12	6	8:2
Single Accelerometer	1,801,306	3	5	8:2

5. Hyperparameter Tuning

Hyperparameter tuning is essential to determine the optimal values of the model settings. The hyperparameter tuning is performed with a grid search on three hyperparameters, which are the optimizer (O), the BiLSTM merge mode (M) and the batch size (B). The optimizers are the Adam optimizer and the Stochastic Gradient Descent with momentum (SGDM), whereas the batch sizes are 8, 16, 32 and 64, and the BiLSTM merge modes are "sum", "multiplication", "average" and "concatenation". The optimal value for each hyperparameter is defined based on the highest test accuracy and the lowest test loss on the UCI-HAR dataset. A summary of the hyperparameter tuning is presented in Table 3.

Table 3. Summary of hyperparameter tuning.

Hyperparameters	Tested Values	Optimal Value
Optimizer (O)	Adam, SGDM	SGDM
BiLSTM Merge Mode (M)	sum, mul, concat, ave	concat
Batch Size (B)	8, 16, 32, 64	64

The results in Table 4 demonstrate that the SGDM does a better job in optimizing the 1D-CNN-BiLSTM model with the ideal batch size of 64 and the concatenation of the outcomes from both LSTM layers. The highest testing accuracy of 95.48%, and the lowest testing loss of 17.62% is accomplished with these hyperparameter settings.

Table 4. Testing accuracy and loss of different optimizers (M = concat, B = 64).

Optimizer (O)	Testing Accuracy (%)	Testing Loss (%)
Adam	95.08	23.20
SGDM	**95.48**	**17.62**

SGDM is a variant of stochastic gradient descent (SGD) where the momentum is incorporated. One reason to integrate the momentum with SGD is that SGD tends to become stuck in the local minimum in the search space. Specifying the value of momentum helps to direct the time series to the correct target by the updates on the previous iterations. Apart from that, momentum also helps to speed up the gradient convergence in proper directions where it brings the time series nearer to the actual derivatives.

Table 5 shows the test accuracy and testing loss of the proposed 1D-CNN-BiLSTM model at different BiLSTM merge modes. In the BiLSTM model, there are two LSTM layers that need to be trained; a merge mode is thus required to integrate the outputs of both layers during the training process. There are four options for the Bi-LSTM merge mode involving sum, multiplication (mul), average (ave) or concatenation (concat). As the name suggests, "sum" and "multiplication" mean that the outputs are totalled or multiplied, respectively. Alternatively, the mean of the outputs is acquired if the "average" mode is chosen, while the outputs are concatenated to the subsequent layer if the "concatenation" mode is applied.

Table 5. Testing Aaccuracy and loss of different BiLSTM merge modes (O = SGDM, B = 64).

Bi-LSTM Merge Modes (M)	Testing Accuracy (%)	Testing Loss (%)
sum	95.05	19.21
mul	93.86	20.47
concat	**95.48**	**17.62**
ave	94.57	19.13

In the experiments, the highest recognition rate is achieved with the "concatenation" setting. This is due to the model having more context and information to learn and preserve from the concatenated outputs of the BiLSTM layers.

Table 6 lists the testing accuracy and testing loss of the proposed 1D-CNN-BiLSTM model with different batch sizes of 8, 16, 32 and 64. It is observed that overall, the testing accuracy of 1D-CNN-BiLSTM rises when B increases. The batch size determines the number of samples that are fed to the network for each iteration during the training process. In this way, it influences the model's learning speed and the steadiness of the learning process.

The most optimal testing accuracy and testing loss are attained with the larger batch size of 64. It is known that the deep neural networks are trained with gradient descent where the computed error from the instances is used to update the weight on each loop. Having a larger batch size results in a more stable model by updating the weight based on

more samples. Additionally, a larger batch size can also help in inhibiting overfitting of the proposed 1D-CNN-BiLSTM model.

Table 6. Testing accuracy and loss of different batch sizes (O = SGDM, M = concat).

Batch Sizes (B)	Testing Accuracy (%)	Testing Loss (%)
8	95.18	18.11
16	95.22	21.88
32	95.05	19.76
64	**95.48**	**17.62**

6. Experimental Results and Analysis

In this section, the experimental results of the proposed 1D-CNN-BiLSTM are first presented. Subsequently, a comparison is conducted between the results of the proposed 1D-CNN-BiLSTM and the existing human activity recognition methods on the UCI-HAR, Motion Sense and Single Accelerometer datasets.

Table 7 presents the performance of the proposed 1D-CNN-BiLSTM in terms of accuracy, precision, recall and F1-score. On the UCI-HAR dataset, the method records 95.48% accuracy and 95.45% F1-score. The 1D-CNN-BiLSTM method yields a relatively lower F1-score of 91.89% on the Motion Sense dataset, probably due to the high inter-class similarity in the sensor data, for instance, upstairs and downstairs, walking and jogging, as well as sitting and standing. Superior performance is observed on the Single Accelerometer dataset with 100% in terms of all evaluation metrics. The performance could be attributable to the high inter-class variance.

Table 7. Experimental results of the 1D-CNN-BiLSTM.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
UCI-HAR	95.48	95.58	95.33	95.45
Motion Sense	94.17	92.33	91.46	91.89
Single Accelerometer	100.00	100.00	100.00	100.00

6.1. Comparative Results with the Existing Works

Table 8 displays the comparative experimental results of the existing methods and 1D-CNN-BiLSTM over the UCI-HAR dataset. In general, the proposed 1D-CNN-BiLSTM model has outstripped the state-of-the-art deep learning methods with an accuracy of 95.48%. The performance of the CNN-based models is also revealing where the recognition rates of RS-1D-CNN [14], FR-DCNN [24], CNN [25] and CNN [11] fall within the range of 94.35% to 95.40%. This is followed by the hybrid LSTM models, where RNN-LSTM [26] and Res-Bidir-LSTM [15] yield an accuracy of 93.89% and 93.6%, respectively.

Succinctly, the proposed 1D-CNN-BiLSTM benefits from having the 1D-CNN layers to extract the salient features from the time series and the BiLSTM layer to connect the long-range relation between the extracted features. The hyperparameter tuning also helps in fitting the proposed 1D-CNN-BiLSTM model with the suitable optimization, merge modes and batch sizes. Thus, each component plays an important role in complementing each other in the model.

Table 8. Comparative results on the UCI-HAR dataset.

Methods	Accuracy (%)
CNN [11]	94.35
RS-1D-CNN [14]	95.40
LSTM [15]	90.80
Bidir-LSTM [15]	91.10
Res-LSTM [15]	91.60
Res-Bidir-LSTM [15]	93.60
CNN-LSTM [16]	92.13
SDAE [17]	95.26
1D-CNN-LSTM [18]	91.04
FR-DCNN [24]	95.27
CNN [25]	95.18
ConvLSTM [26]	92.24
RNN-LSTM [26]	93.89
RNN [27]	95.03
1D-CNN-BiLSTM	**95.48**

Table 9 records the comparative experimental results of the existing methods and 1D-CNN-BiLSTM on the Motion Sense dataset. It can be summarized that the accuracy of deep learning techniques is sorted from low to high as 79.86%, 82.50%, 83.30%, 85.59%, 85.75%, 89.00% and 89.08% of BERT [28], CAE [29], multi-task self-supervision CNN [30], DeepConvLSTM [31], CNN [32], DCNN [33] and CPC [34], in ascending order.

Table 9. Comparative results on the motion sense dataset.

Methods	Accuracy (%)
BERT [28]	79.86
CAE [29]	82.50
Multi-task self-supervision CNN [30]	83.30
DeepConvLSTM [31]	85.59
CNN [32]	85.75
DCNN [33]	89.00
CPC [34]	89.08
1D-CNN-BiLSTM	**94.17**

In comparison with the existing methods, the proposed 1D-CNN-BiLSTM shows a superior recognition rate of 94.17% on the Motion Sense dataset. The proposed 1D-CNN-BiLSTM reveals that the performance can be magnified through the cooperation between the 1D-CNN layers and the Bi-LSTM layer. Both types of layers contribute to the effective human activity recognition where the 1D-CNN layers collect the discriminative features, whereas the BiLSTM layer deduces the temporal information in two directions from the time series data.

Table 10 presents the comparative experimental results of the state of the art and 1D-CNN-BiLSTM on the Single Accelerometer dataset. It can be concluded that the deep learning approaches show more reliable results on the Single Accelerometer dataset than the machine learning approaches in terms of accuracy. One of the major reasons is that the Single Accelerometer dataset contains the greatest number of samples among the datasets, which is up to 1,801,306 instances. As the scale of the instances increases drastically, a deep learning approach is a better alternative as it has the ability to autonomously learn the features given the training data. The experimental results demonstrate that the proposed 1D-CNN-BiLSTM model records the highest accuracy of 100% despite the huge sample size.

Table 10. Comparative results on the single accelerometer dataset.

Methods	Accuracy (%)
SDAE [17]	89.99
CNN [19]	91.77
ConvLSTM [19]	92.29
3D-CNN by ConvLSTM [19]	93.69
Random Forest [35]	88.00
SVM [36]	80.00
Random Forest [36]	94.00
Deep LSTM [37]	91.34
1D-CNN-BiLSTM	**100.00**

6.2. Confusion Matrices

The confusion matrix of the proposed 1D-CNN-BiLSTM model on the UCI-HAR dataset is depicted in Figure 2. Note that the sitting class is often mistakenly recognized for the standing and laying classes due to these activity classes being static, hence producing similar sensor time series. Apart from that, the walking downstairs, walking upstairs, and walking classes are also frequently misclassified to each other as they involve relatively identical limb movements.

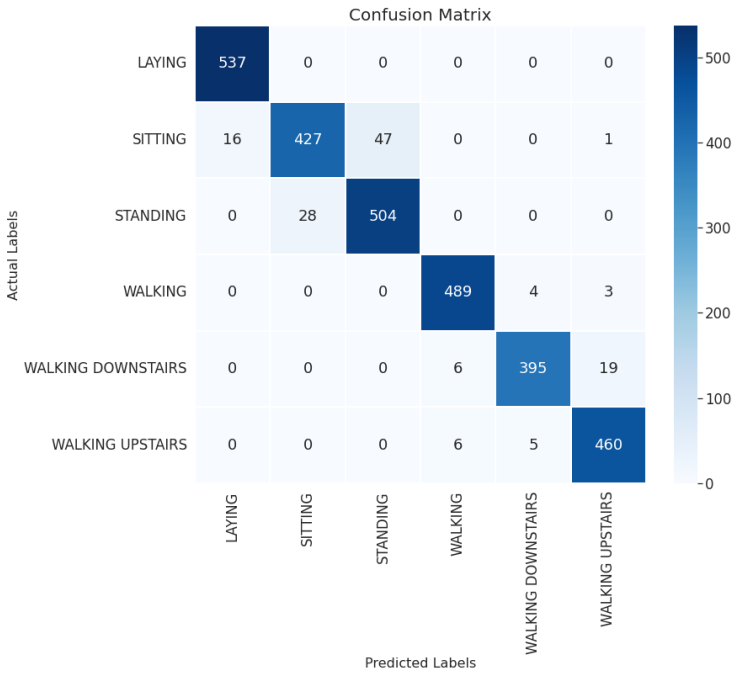

Figure 2. The confusion matrix of the 1D-CNN-BiLSTM model on the UCI-HAR dataset.

Figure 3 illustrates the confusion matrix of the 1D-CNN-BiLSTM model on the Motion Sense dataset. Similar trends are observed on the Motion Sense dataset where the majority of the misclassifications occur among the downstairs, upstairs, walking and jogging classes. These activities exhibit similar limb movements with minor deviations that the wearable sensors can hardly capture, thus slightly deteriorating the recognition performance.

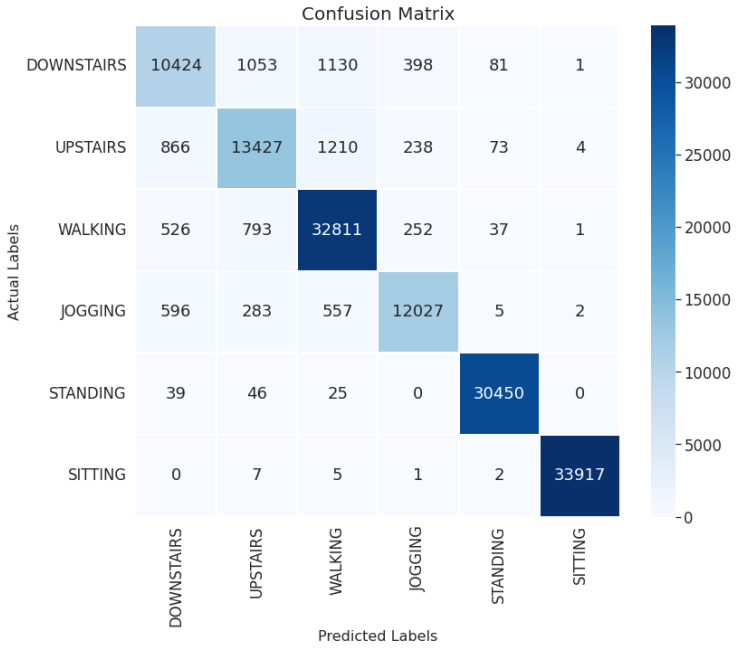

Figure 3. The confusion matrix of the 1D-CNN-BiLSTM model on the Motion Sense dataset.

On the Single Accelerometer dataset, all testing samples are correctly classified as displayed in Figure 4. The activity classes of this dataset are comparatively well separated, where the highly confounding classes, such as walking upstairs and walking downstairs are grouped into one activity class. The discriminative capability of the 1D-CNN-BiLSTM model and the inter-class disparities in the accelerometer sensor data collectively contribute to the spectacular performance on the dataset.

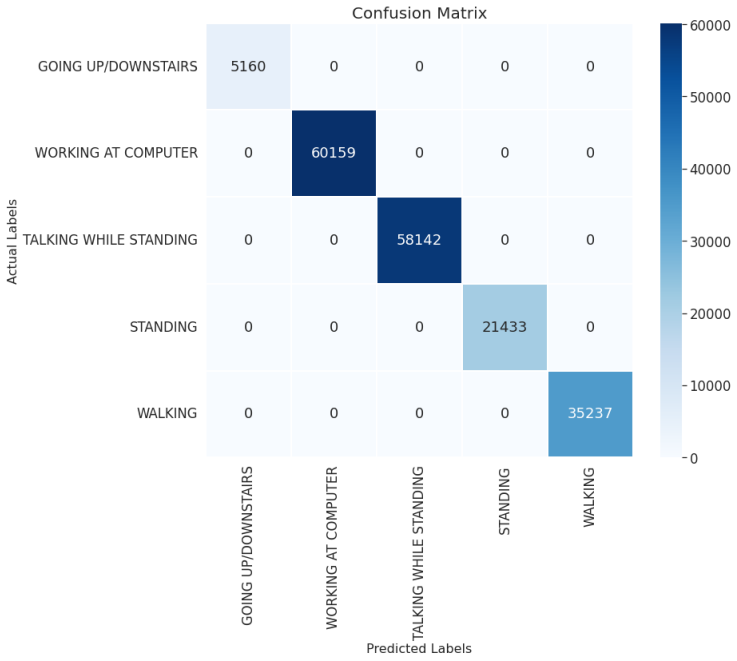

Figure 4. The confusion matrix of the 1D-CNN-BiLSTM model on the Single Accelerometer dataset.

7. Conclusions

This paper presents a hybrid deep learning model, referred to as the 1D-CNN-BiLSTM, for sensor-based human activity recognition. The 1D-CNN-BiLSTM model leverages the 1D-CNN layers to encode the sensor data into the features at different granularity.

Subsequently, the BiLSTM layer encodes the long-range dependencies in the features to preserve the class specific properties. The performance of the proposed 1D-CNN-BiLSTM model is compared using three datasets, namely the UCI-HAR dataset, the Motion Sense dataset, and the Single Accelerometer dataset. Compared with the existing human activity recognition methods, 1D-CNN-BiLSTM outshines the state-of-the-art methods , with a recognition rate of 95.48% on the UCI-HAR dataset, 94.17% on the Motion Sense dataset and 100% on the Single Accelerometer dataset. The proposed 1D-CNN-BiLSTM provides significant improvements in human activity recognition as the deep learning models are able to adapt to the new time series that are constantly being fed, learn and reason everything on its own for the desired outcome. In future work, several enhancements could be performed, such as oversampling for class imbalanced problems and data augmentation for improved model generalization capability.

Author Contributions: Conceptualization, Y.J.L. and C.P.L.; methodology, Y.J.L. and C.P.L.; software, Y.J.L. and C.P.L.; validation, Y.J.L. and C.P.L.; formal analysis, Y.J.L.; investigation, Y.J.L.; resources, Y.J.L.; data curation, Y.J.L. and C.P.L.; writing—original draft preparation, Y.J.L.; writing—review and editing, C.P.L. and K.M.L.; visualization, Y.J.L. and C.P.L.; supervision, C.P.L. and K.M.L.; project administration, C.P.L.; funding acquisition, C.P.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research in this work was supported by the Fundamental Research Grant Scheme of the Ministry of Higher Education under award number FRGS/1/2021/ICT02/MMU/02/4 and Multimedia University Internal Research Grant with award number MMUI/220021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ranasinghe, S.; Al Machot, F.; Mayr, H.C. A review on applications of activity recognition systems with regard to performance and evaluation. *Int. J. Distrib. Sens. Netw.* **2016**, *12*, 1550147716665520. [CrossRef]
2. Tan, P.S.; Lim, K.M.; Lee, C.P. Human action recognition with sparse autoencoder and histogram of oriented gradients. In Proceedings of the 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 26–27 September 2020; pp. 1–5.
3. Irvine, N.; Nugent, C.; Zhang, S.; Wang, H.; Ng, W.W. Neural network ensembles for sensor-based human activity recognition within smart environments. *Sensors* **2019**, *20*, 216. [CrossRef] [PubMed]
4. Neili Boualia, S.; Essoukri Ben Amara, N. Deep full-body HPE for activity recognition from RGB frames only. *Informatics* **2021**, *8*, 2. [CrossRef]
5. Dobbins, C.; Rawassizadeh, R. Towards clustering of mobile and smartwatch accelerometer data for physical activity recognition. *Informatics* **2018**, *5*, 29. [CrossRef]
6. Lee, C.P.; Lim, K.M.; Woon, W.L. Statistical and entropy based multi purpose human motion analysis. In Proceedings of the 2010 2nd International Conference on Signal Processing Systems, Dalian, China, 5–7 July 2010; Volume 1, p. V1-734.
7. Saez, Y.; Baldominos, A.; Isasi, P. A comparison study of classifier algorithms for cross-person physical activity recognition. *Sensors* **2016**, *17*, 66. [CrossRef] [PubMed]
8. Dirgová Luptáková, I.; Kubovčík, M.; Pospíchal, J. Wearable sensor-based human activity recognition with transformer model. *Sensors* **2022**, *22*, 1911. [CrossRef]
9. Zhang, S.; Li, Y.; Zhang, S.; Shahabi, F.; Xia, S.; Deng, Y.; Alshurafa, N. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors* **2022**, *22*, 1476. [CrossRef]
10. Murad, A.; Pyun, J.Y. Deep recurrent neural networks for human activity recognition. *Sensors* **2017**, *17*, 2556. [CrossRef]
11. Ignatov, A. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Appl. Soft Comput.* **2018**, *62*, 915–922. [CrossRef]
12. Moya Rueda, F.; Grzeszick, R.; Fink, G.A.; Feldhorst, S.; Ten Hompel, M. Convolutional neural networks for human activity recognition using body-worn sensors. *Informatics* **2018**, *5*, 26. [CrossRef]
13. Ferrari, A.; Micucci, D.; Mobilio, M.; Napoletano, P. Hand-crafted features vs residual networks for human activities recognition using accelerometer. In Proceedings of the 2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT), Ancona, Italy, 19–21 June 2019; pp. 153–156.

14. Ragab, M.G.; Abdulkadir, S.J.; Aziz, N. Random search one dimensional CNN for human activity recognition. In Proceedings of the 2020 International Conference on Computational Intelligence (ICCI), Bandar Seri Iskandar, Malaysia, 8–9 October 2020; pp. 86–91.
15. Zhao, Y.; Yang, R.; Chevalier, G.; Xu, X.; Zhang, Z. Deep residual bidir-LSTM for human activity recognition using wearable sensors. *Math. Probl. Eng.* **2018**, *2018*, 7316954. [CrossRef]
16. Mutegeki, R.; Han, D.S. A CNN-LSTM approach to human activity recognition. In Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 19–21 February 2020; pp. 362–366.
17. Ni, Q.; Fan, Z.; Zhang, L.; Nugent, C.D.; Cleland, I.; Zhang, Y.; Zhou, N. Leveraging wearable sensors for human daily activity recognition with stacked denoising autoencoders. *Sensors* **2020**, *20*, 5114. [CrossRef] [PubMed]
18. Goh, J.X.; Lim, K.M.; Lee, C.P. 1D Convolutional Neural Network with Long Short-Term Memory for Human Activity Recognition. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 13–15 September 2021; pp. 1–6.
19. Erdaş, Ç.B.; Güney, S. Human activity recognition by using different deep learning approaches for wearable sensors. *Neural Process. Lett.* **2021**, *53*, 1795–1809. [CrossRef]
20. Hamad, R.A.; Yang, L.; Woo, W.L.; Wei, B. Joint learning of temporal models to handle imbalanced data for human activity recognition. *Appl. Sci.* **2020**, *10*, 5293. [CrossRef]
21. Anguita, D.; Ghio, A.; Oneto, L.; Parra Perez, X.; Reyes Ortiz, J.L. A public domain dataset for human activity recognition using smartphones. In Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 24–26 April 2013; pp. 437–442.
22. Malekzadeh, M.; Clegg, R.G.; Cavallaro, A.; Haddadi, H. Mobile sensor data anonymization. In Proceedings of the International Conference on Internet of Things Design and Implementation, Montreal, QC, Canada, 15–18 April 2019; pp. 49–58.
23. Casale, P.; Pujol, O.; Radeva, P. Personalization and user verification in wearable systems using biometric walking patterns. *Pers. Ubiquitous Comput.* **2012**, *16*, 563–580. [CrossRef]
24. Qi, W.; Su, H.; Yang, C.; Ferrigno, G.; De Momi, E.; Aliverti, A. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. *Sensors* **2019**, *19*, 3731. [CrossRef]
25. Jiang, W.; Yin, Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 1307–1310.
26. Roobini, M.S.; Naomi, M.J.F. Smartphone sensor based human activity recognition using deep learning models. *Int. J. Recent Technol. Eng.* **2019**, *8*, 2740–2748.
27. Inoue, M.; Inoue, S.; Nishida, T. Deep recurrent neural network for mobile human activity recognition with high throughput. *Artif. Life Robot.* **2018**, *23*, 173–185. [CrossRef]
28. Haresamudram, H.; Beedu, A.; Agrawal, V.; Grady, P.L.; Essa, I.; Hoffman, J.; Plötz, T. Masked reconstruction based self-supervision for human activity recognition. In Proceedings of the 2020 International Symposium on Wearable Computers, Virtual Event, 12–16 September 2020; pp. 45–49.
29. Haresamudram, H.; Anderson, D.V.; Plötz, T. On the role of features in human activity recognition. In Proceedings of the 23rd International Symposium on Wearable Computers, London, UK, 9–13 September 2019; pp. 78–88.
30. Saeed, A.; Ozcelebi, T.; Lukkien, J. Multi-task self-supervised learning for human activity detection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–30. [CrossRef]
31. Ordóñez, F.J.; Roggen, D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115. [CrossRef]
32. Ferreira, J.; Carvalho, E.; Ferreira, B.V.; de Souza, C.; Suhara, Y.; Pentland, A.; Pessin, G. Driver behavior profiling: An investigation with different smartphone sensors and machine learning. *PLoS ONE* **2017**, *12*, e0174959.
33. Ilisei, D.; Suciu, D.M. Human-activity recognition with smartphone sensors. In Proceedings of the OTM Confederated International Conferences "On the Move to Meaningful Internet Systems", Rhodes, Greece, 21–25 October 2019; pp. 179–188.
34. Haresamudram, H.; Essa, I.; Plötz, T. Contrastive predictive coding for human activity recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–26. [CrossRef]
35. Erdaş, Ç.B.; Atasoy, I.; Açıcı, K.; Oğul, H. Integrating features for accelerometer-based activity recognition. *Procedia Comput. Sci.* **2016**, *98*, 522–527. [CrossRef]
36. Hossain, T.; Inoue, S. A Comparative study on missing data handling using machine learning for human activity recognition. In Proceedings of the 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 30 May–2 June 2019; pp. 124–129.
37. Güney, S.; Erdaş, Ç.B. A deep LSTM approach for activity recognition. In Proceedings of the 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 1–3 July 2019; pp. 294–297.