Vol 6 No 2 (2024) *e*-ISSN: 2682-8383

Journal of Engineering Technology and Applied Physics

Temporal Convolutional Recurrent Neural Network for Elderly Activity Recognition

Ng Jia Hui[1], Pang Ying Han[1,*], Sarmela Raja Sekaran[1], Ooi Shih Yin[1] and Lillian Yee Kiaw Wang[2]
[1]*Faculty of Information Science and Technology, Multimedia University, Ayer Keroh, Melaka, Malaysia.*
[2]*School of Information Technology, Monash University, Selangor, Malaysia.*
**Corresponding Author*: yhpang@mmu.edu.my, *ORCiD*: 0000-0002-3781-6623
https://doi.org/10.33093/jetap.2024.6.2.12
Manuscript Received: 18 April 2024, Accepted: 10 June 2024, Published: 15 September2024

***Abstract* — Research on smartphone-based human activity recognition (HAR) is prevalent in the field of healthcare, especially for elderly activity monitoring. Researchers usually propose to use of accelerometers, gyroscopes or magnetometers that are equipped in smartphones as an individual sensing modality for human activity recognition. However, any of these alone is limited in capturing comprehensive movement information for accurate human activity analysis. Thus, we propose a smartphone-based HAR approach by leveraging the inertial signals captured by these three sensors to classify human activities. These heterogeneous sensors deliver information on various aspects of nature, motion and orientation, offering a richer set of features for more accurate representations of the activities. Hence, a deep learning approach that amalgamates long short-term memory (LSTM) in temporal convolutional network (TCN) is proposed. We use independent temporal convolutional networks, coined as temporal convolutional streams, to independently analyse the temporal data of each sensing modality. We name this architecture multi-stream TC-LSTM. The performance of multi-stream TC-LSTM is assessed on the self-collected elderly activity database. Empirical results exhibit that multi-stream TC-LSTM outperforms the existing machine learning and deep learning models, with an F1 score of 98.3%.**

Keywords*— Elderly Activity Recognition, Multi-Stream, Recurrent Neural Network, Deep Learning, Temporal Convolutional.

I. INTRODUCTION

The global elderly population continues to grow at a rapid rate. Therefore, developing human activity recognition (HAR) systems that are specifically tailored for elderly population is significant. These systems are crucial for their healthcare and safety needs. In recent years, AI-oriented tools have been widely employed in HAR research [1–3]. In other words, a wide range of machine learning and artificial intelligence techniques are explored to empower the capabilities of HAR systems [4]. By interpreting and understanding the behaviours of individuals, the applications provide support for the individuals' needs and preferences. Generally, human activities are categorized into: (1) basic activities and (2) complex activities. The former activities are those actions that are relatively simple, such as walking, running and standing. On the other hand, complex activities are those actions that are more diverse and context-dependent, involving a combination of basic activities. Examples of complex activities are eating a meal, drinking a cup of water, cooking a meal, driving a car, playing badminton, etc. Machine learning and deep learning algorithms have been extensively employed for an efficient HAR system to analyse the data and extract informative features [5–7].

Nowadays, smartphones are an integral part of our lives. Almost everyone owns a smartphone. Thus, this ubiquitous presence of smartphones makes them a convenient platform for HAR. Researchers propose to use of accelerometers, gyroscopes or magnetometers that are equipped in smartphones as an individual sensing modality for HAR. However, any of these alone is limited in capturing comprehensive movement information for accurate human activity analysis. Thus, we propose a smartphone-based HAR approach by leveraging the inertial signals captured by these three sensors to classify human activities. These heterogeneous sensors deliver information on various aspects of nature, motion and orientation, offering a richer set of features for more accurate representations of the activities.

Hence, we present a deep learning architecture – integrating a temporal convolutional network and a recurrent neural network for human activity analysis

Published by MMU PRESS. URL: https://journals.mmupress.com/index.php/jetap/index

and classification. Specifically, Temporal Convolutional Network (TCN) and long short-term memory (LSTM) architecture are incorporated to achieve improved performance in activity recognition. The key objective of the proposed model, named multi-stream TC-LSTM, is to directly process raw inertial data without preprocessing or supplemental domain knowledge and effectively capture relevant temporal patterns of the data. Multi-stream TC-LSTM leverages the strengths of TCN and LSTM for reliable human activity recognition.

The overview of the proposed architecture is illustrated in Fig. 1. In this model, low-level features are computed using TCN that encodes spatial-temporal information. Next, these low-level features are further input into a classifier to capture high-level temporal information using LSTM. The hierarchical extraction of these two levels of information facilitates more accurate predictions. Unlike other approaches [8–11], our proposed model utilizes a distant TCN for processing each individual sensor data. Each sensor instance of the convolution process is referred to as a convolutional stream, thereby forming a Multi-stream TC-LSTM architecture. This implementation enhances feature extraction by focusing on one specific sensor rather than on all sensors simultaneously. Moreover, the proposed architecture is adaptable to new sensor configurations since the convolutional streams are easily removed, added or modified. The adoption of LSTM in this model is due to its excellence at modelling sequential data. LSTM process utilizes past information. Contrasting recurrent neural networks (RNNs) that are prone to the vanishing gradient issue, LSTM addresses this problem via their unique memory cell structure which enables effectively capturing information over longer time horizons. This feature is significant for learning temporal patterns.

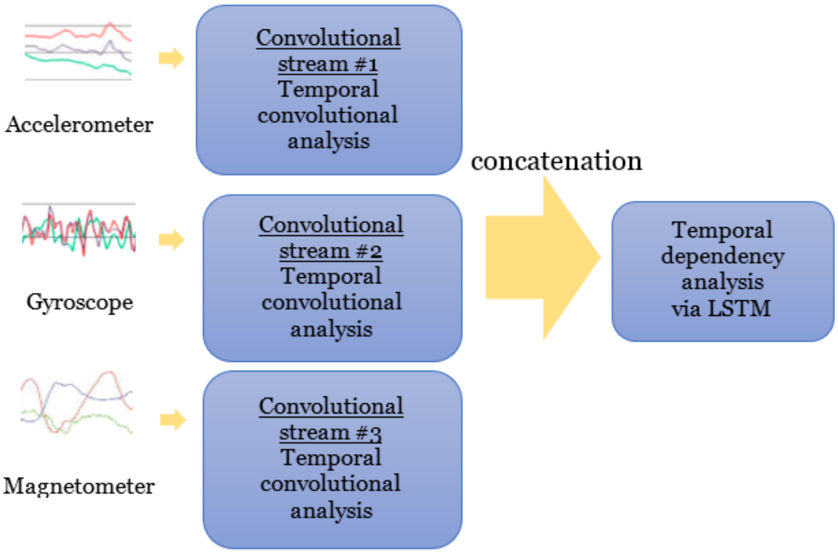

Fig. 1. The overview of the proposed Multi-stream TC-LSTM architecture.

The contributions of this paper are:

1. An end-to-end human activity recognition system is developed. The proposed system directly works on the captured raw inertial data. There is no preprocessing, nor any prior domain knowledge is required.

2. By utilizing individual temporal convolutions, the proposed Multi-stream TC-LSTM independently processes and analyses data from different sensor types. This allows customized convolutional processes tailored to befit specific data properties or features of each sensor, extracting more meaningful information from the sensor data.

3. Since convolutional streams are completely independent to each other, this modularity facilitates straightforward integration of additional sensor inputs by just adding new convolutional streams for the data of the new sensors. Similarly, if a sensor is removed, the corresponding convolutional stream can be simply removed from the system.

II. RELATED WORK

Deep learning models have demonstrated outstanding performance across diverse applications, ranging from speech recognition to video classification. Their capability to extract informative patterns and dependencies from raw data makes them a preferred feature analysis approach for sensor-based human recognition. Convolutional Neural Network (CNN)-based deep learning models have been extensively utilized for HAR [5, 9, 10]. Cho *et al.* proposed a CNN-based deep model with bidirectional LSTM and CNN to recognize human activity [5]. Two publicly available databases which are WISDM and UCI-HAR were used for the model performance assessment. The proposed CNN-based model outperforms the existing approaches in smartphone-based HAR. Furthermore, Kim *et al.* introduced a wearable IMU to evaluate patients through BBS (Berg balance scale) for balance assessment [12]. The proposed stacking ensemble model contains a dense layer comprising two one-dimensional-CNN heads and a GRU head. The obtained empirical results demonstrated that the model's computational complexity and performance were improved. The authors highlighted that their proposed algorithm is potential to a wearable healthcare device for users to understand their balance ability and probability of failing, helpful for falling prevention.

Recurrent models are popular for time series data classification, and human activity recognition is a kind of time series classification [13]. In other words, recurrent models can extract temporal dependencies from inertial signals and these temporal features are crucial in comprehending the progression of motion activities over time. A lightweight Recurrent Neural Network (RNN) is proposed on a low-power microcontroller for HAR [14]. The experimental results demonstrated the feasibility of the RNN model in the constrained-resource scenario. The proposed system attained a promising recognition performance while maintaining low memory consumption.

A variant of RNN, i.e. bidirectional LSTM, was utilized to classify and recognize human activity [15]. In this work, a robust HAR model is developed by combining Deep SqueezeNet and bidirectional LSTM. Furthermore, an improved flower pollination optimization algorithm, known as IFPOA, was adopted for bidirectional LSTM's hyperparameter optimization. The integration of SqueezeNet and bidirectional LSTM enables the model to extract short- and long-term dependencies from the data. Nevertheless, when dealing with long data sequences, recurrent models are computationally expensive, causing high memory consumption, especially when dealing with long sequences as well as large batch sizes.

In recent years, the employment of Temporal Convolutional Networks (TCNs) has been widespread in smartphone-based human activity recognition. The key factors are their efficient parallel processing and capability to capture long-term dependencies while requiring lower memory. Sarmela *et al.* proposed a lightweight TCN-based deep learning model for HAR [16]. The model extracts features from inertial data through parallel Convolutional Heads. Besides that, dilated causal convolutions and residual connections are incorporated into the model to preserve the longer-term dependencies of the data. The model's performance was assessed on three databases and promising performances were obtained. Linh and Bach proposed a temporal convolutional network with Hidden Markov Chain for human activity recognition post-processing [17]. In the paper, an algorithm was presented to produce huge noisy data for weight initialization in the model. In the post-processing stage, the authors proposed a technique to smooth the model's prediction. A promising results are obtained.

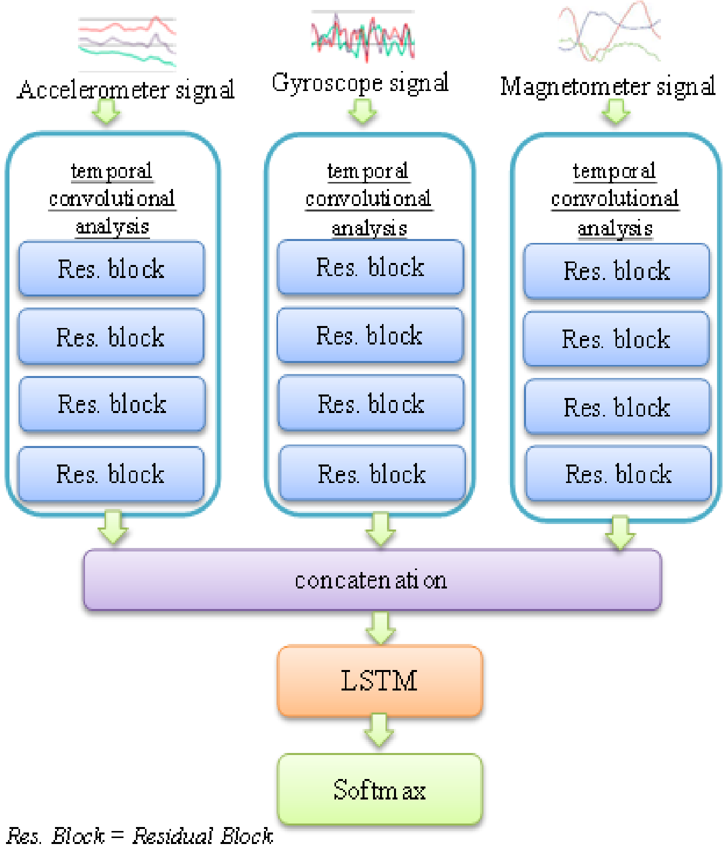

Fig. 2. The proposed system architecture.

III. PROPOSED METHOD DESIGN

There are multiple sensors embedded in smartphones. Any of these sensors alone is limited in capturing inclusive movement information for accurate human activity analysis and recognition. In this paper, we propose a smartphone-based HAR approach by leveraging the inertial signals captured by three sensors, i.e. accelerometer, gyroscope or magnetometer, to classify human activities. The adoption of multiple heterogeneous sensors enables the collection of diverse information on different aspects of nature, motion and orientation. This rich information is crucial to represent the underlying dynamics of activity data. In the proposed model, a temporal convolutional (TC) network and long short-term memory (LSTM) are integrated. We name this architecture multi-stream TC-LSTM.

Figure 2 illustrates the multi-stream TC-LSTM architecture for multiple sensing modalities analysis. Firstly, the captured sensor data is processed. Unlike other approaches [8–11], independent temporal convolutional analysis is performed on each sensor data. Each sensor instance of the convolution process is coined as a convolutional stream. In other words, multiple temporal convolutional streams are conducted to independently process the temporal data of different sensing modalities captured by multiple sensors. This independent sensor-specific data analysis enables a more thorough extraction of diverse information from different sensing modalities. In this study, TCN is adopted for temporal convolutional analysis. The primary reasons for applying TCN are (1) TCN performs well for sequential data [18], and (2) causal convolution is implemented in TCN in such a way that the convolutional operation does not take future values as inputs, preventing information leakage from the future to the past.

In the proposed model, each TCN stream is constituted by residual blocks. In each residual block, there are temporal convolutions, batch normalizations, ReLU activation functions and spatial dropouts, as illustrated in Fig. 3. The stacking of multiple residual blocks in each TCN stream empowers the network to effectively capture the hierarchical features of the data sequences. Since the sequential data passes through consecutive layers, the network can abstract and represent the complex patterns progressively. In this work, dilated temporal convolutions are performed to enlarge the receptive fields, enabling long-term dependency analysis to extract data characteristics and relationships across a temporal span. From Fig. 3, we can notice that there is a 1×1 convolution in a residual block. This is called residual connection and is a key feature of a residual block. The residual connection is vital as it allows the gradient to pass directly through the block, even when the gradients become very small values due to reiterated convolutions and activations. This feature helps address the vanishing gradient problem and facilitates the training of a deeper network.

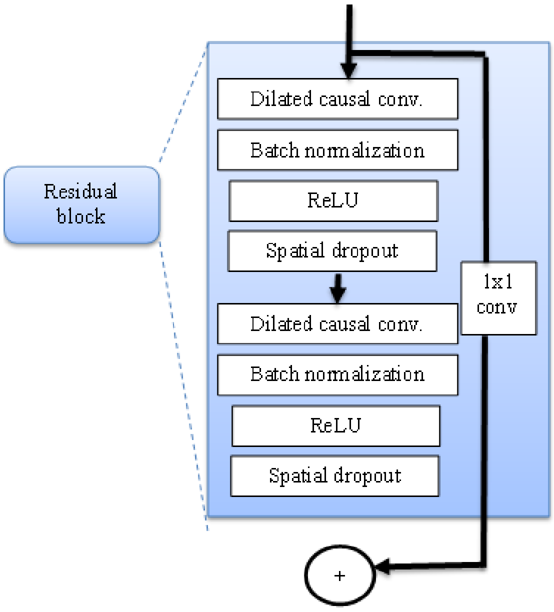

Fig. 3. Residual block.

In order to capture long-term patterns of the inertial signals, we apply one-dimensional dilated convolutions in TCNs for larger receptive fields to effectively capture information from distant time paces in the sequence data, as illustrated in Fig. 4. The dilated causal convolution process can be formulated as follows,

$$x_l^t = h\left(\sum_{k=0}^{K-1} w_l^k x_{(l-1)}^{(t-(k\times d))} + b_l\right) \quad (1)$$

, where x_l^t is the output of the neuron at t position in the lth layer; K is the length of the kernel; x_l^t is the weight of k position; d is the dilation factor of the convolution; b_l is a bias. In this work, the Rectified Linear Units (ReLU) activation function, i.e. $h(x) = \max(0, x)$, so that the model can capture data nonlinearities to handle complex interactions. Besides, the implementation of batch normalisations speeds and stabilises model training. Spatial dropout helps fix vanishing gradients by regularising. Spatial dropout trains the network to learn more robust and generalisable properties by randomly losing neurons.

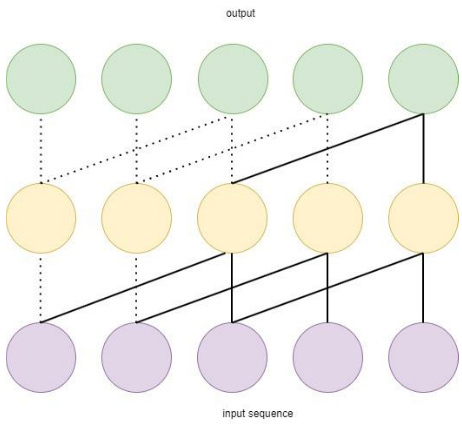

Fig. 4. Dilated causal convolution with two layers with kernel length 2 and dilated rate 2.

After the temporal convolution analysis, the generated feature maps of each stream are concatenated and further processed by LSTM to improve the temporal abstraction of the inertial signals. This amalgamation of TCN and LSTM imparts a multi-level abstraction of the inertial signals in such a way that TCN produces initial abstraction at fixed temporal scales, and LSTM refines the abstraction by adjusting its gating mechanisms and memory cells to learn temporal variations in the data which are crucial to exhibit assorted movement dynamics. Furthermore, unlike traditional recurrent neural networks (RNNs) that suffer from the vanishing gradient problem, LSTM addresses this issue via their unique memory cell structure which enables effectively capturing information over longer time horizons. This feature is significant for learning temporal patterns in motion data.

An LSTM network comprises an input layer, hidden layers and an output layer. Memory cells contained in the hidden layers and each cell comprises 3 gates to maintain and regulate the cell state (forget gate, input gate and output gate). The structure of a memory cell is presented in Fig. 5.

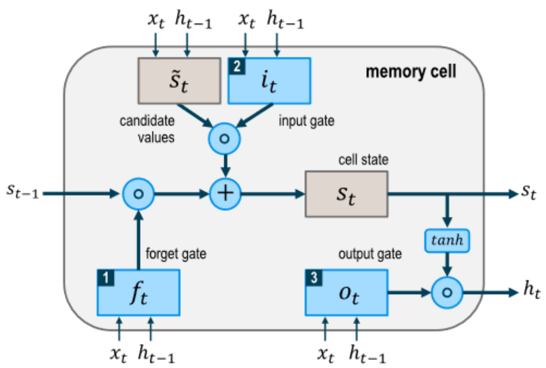

Fig. 5. LSTM memory cell structure (source: Giang *et al.*, 2022).

IV. EXPERIMENT RESULT AND DISCUSSION

The effectiveness of the proposed framework is evaluated on our self-collected elderly activity recognition dataset. In the literature, there are relatively fewer studies on elderly HAR compared to activity recognition focused on adults. Hence, in this paper, elderly activity recognition is focused. There are 13 volunteers in the age range of 60 years old and above. Among the participants, there are eight females and five males. These elderly people can move and behave normally, though some have age-related conditions such as high cholesterol, high blood pressure and/or high blood sugar levels.

The data collection was conducted at the participants' homes and in the home yards. During the data collection, the volunteers placed a Samsung Galaxy A52s 5G smartphone in their left front pocket. Each participant was required to perform six basic activities: standing, sitting, laying, brisk walking and stair climbing. Each activity was recorded for a total duration of 2 minutes. This could be split into two one-minute sessions performed at separate times upon the request of the elderly. At the sampling rate of 20ms, acceleration signals were acquired. A sliding window (with an overlap rate of 50%) was applied to segment the inertial data sequence with a window size of 128 data points. Each data sample (also referred to as window) is represented by the *x*, *y* and *z* coordinates of the acceleration signal data. In total, the database contains 3762 data samples.

After the data sampling comprising data collection and segmentation, the data sample was preprocessed to scale the features within the range between 0 and 1. This process is to ensure the equal contribution of all features to the data analysis, preventing features with larger scales from dominating. The collected dataset was then randomly split into three sets, with 70% for model training data, 15% for model validation and 15% for model testing.

A. *Performance Metrics*

In this study, various performance metrics are considered to evaluate the performance of multi-stream TC-LSTM:

$$Precision = \frac{T_p}{T_p + F_p} \tag{2}$$

$$Recall = \frac{T_p}{T_p + F_n} \tag{3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

, where T_p is a true positive, F_p is a false positive, and F_n is a false negative.

B. *Analysis of Model Hyperparameters*

Hyperparameters are essential in developing and optimizing deep learning models. They directly impact the capability of a deep learning model to learn data patterns and generate accurate predictions. In this section, we examine the impact of specific hyperparameters on the performance of the proposed multi-stream TC-LSTM. The effect of dilation factors on activity recognition performance is analysed and recorded in Table 1. Based on the empirical results, we can observe that the performance of the proposed model slightly declines as the dilation rates increase. While the dilation factor exerts a certain influence on the model's performance, the influence is minor. However, dilation rates significantly influence the model's complexity. A greater number of model parameters are generated with higher dilation rates. Excessively complex models could trigger model overfitting, wherein the models capture random fluctuations and noise present in the training data. This consequence is detrimental to the model's overall performance, resulting in unreliable predictions. This finding is illustrated in Table I by the decrease in F1 score when dilation rates are extended up to 32.

Table I. Performance of different dilation factors in the proposed model.

Dilation Rate	No of Parameters	Precision	Recall	F1 Score
1,2,4	6059538	0.9836	0.9845	0.9839
1,2,4,8	8173074	0.9738	0.9679	0.9702
1,2,4,8,16	10286610	0.9778	0.9789	0.9782
1,2,4,8,16,32	12400146	0.9708	0.9714	0.97

Next, the effect of dropout factors on the performance of activity recognition is studied. Dropout imposes regularization into the model to counter overfitting. However, excessive use of dropout can hinder the model from learning data patterns. Table II records the performance of different dropout factors of the proposed model. From the results, we can observe that a dropout rate of 0.05 attains the best performance with a high F1 score. We can deduce that with this dropout rate value, the proposed model is able to maintain its capability of capturing informative data patterns while preventing overfitting. When the dropout rates increase, there is a noticeable degradation in the recognition performance. Performance degradation could potentially arise due to the excessive reduction of neuron activations during training (associated with higher dropout rates), leading to underfitting. The model fails to learn the underlying patterns in the data.

Table II. Performance of different dropout factors in the proposed model.

Dropout Rate	Precision	Recall	F1 Score
0.05	0.9892	0.9885	0.9889
0.1	0.9803	0.9795	0.979
0.2	0.9396	0.9368	0.9358

C. Performance Analysis

In this section, the performance analysis of the proposed model is conducted. The recognition performance of each activity class is summarized in Table III. Furthermore, we also present the performances of multi-stream Temporal Convolutional Network (i.e. multi-stream TC). This analysis is to elucidate the contribution and significance of the LSTM component within the proposed multi-stream TC-LSTM. In other words, by excluding and scrutinizing the function of LSTM, we examine the performances of the proposed multi-stream TC-LSTM (i.e. with LSTM) and multi-stream TC (i.e. without LSTM).

The empirical results demonstrate that the proposed multi-stream TC-LSTM achieves better recognition performance with an F1 score of 98.3%; whereas when the LSTM component is excluded from the architecture, the performance of the model degrades with approximately 4% in the F1 score. This analysis suggests that the LSTM component contributes considerably in capturing intricate temporal patterns from human activity signals. The recurrent nature of LSTM facilitates an effective sequential dependency modelling for the inertial data and enables the model to comprehend the underlying dynamics for characterizing the activities.

Table III. Performance of the proposed Multi-stream TC-LSTM and Multi-stream TC (without LSTM).

Activity	Precision	Recall	F1 Score
Multi-stream TC-LSTM			
Brisk Walking	0.94	0.99	0.96
Climbing Stairs	1.0	0.99	1.0
Laying	1.0	1.0	1.0
Sitting	1.0	1.0	1.0
Standing	1.0	0.99	0.99
Walking	0.96	0.94	0.95
Mean	0.983	0.985	0.983
Multi-stream TC			
Brisk Walking	0.79	0.86	0.82
Climbing Stairs	1.0	1.0	1.0
Laying	1.0	1.0	1.0
Sitting	1.0	1.0	1.0
Standing	1.0	1.0	1.0
Walking	0.88	0.82	0.85
Mean	0.945	0.947	0.945

Additionally, the confusion matrix of multi-stream TC-LSTM is also depicted in Fig. 6. We can notice that five walking samples have been misclassified into brisk walking class, and one sample of brisk walking, climbing stairs and standing, respectively, have been misclassified into walking class. This could be because of the similar inertial patterns between walking and brisk walking, especially on the gyroscope signals, as shown in Fig. 7. Besides that, we also observe that the proposed model achieves good recognition performance in distinguishing passive activities such as laying, sitting and standing. This may result from the distinct nature of the axial signals (blue, green and orange coloured) of the accelerometer and gyroscope of these classes, see Fig. 8.

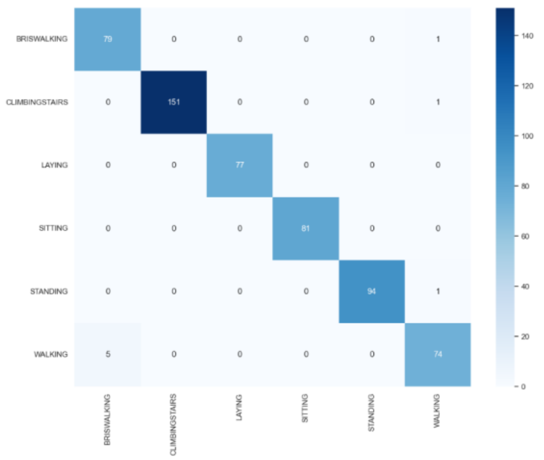

Fig. 6. Confusion matrix of the Multi-stream TC-LSTM.

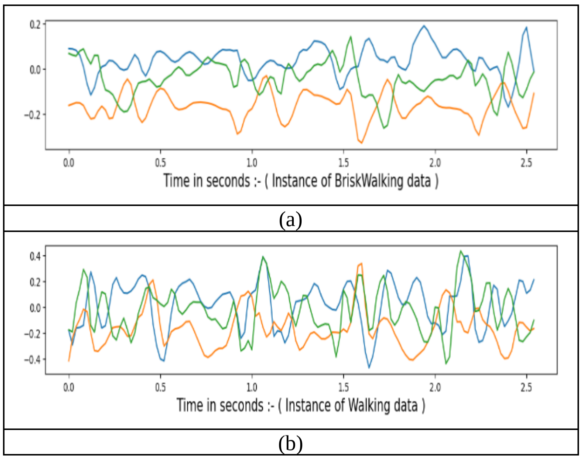

Fig. 7. Gyroscope signal patterns of a (a) brisk walking sample and (b) walking sample.

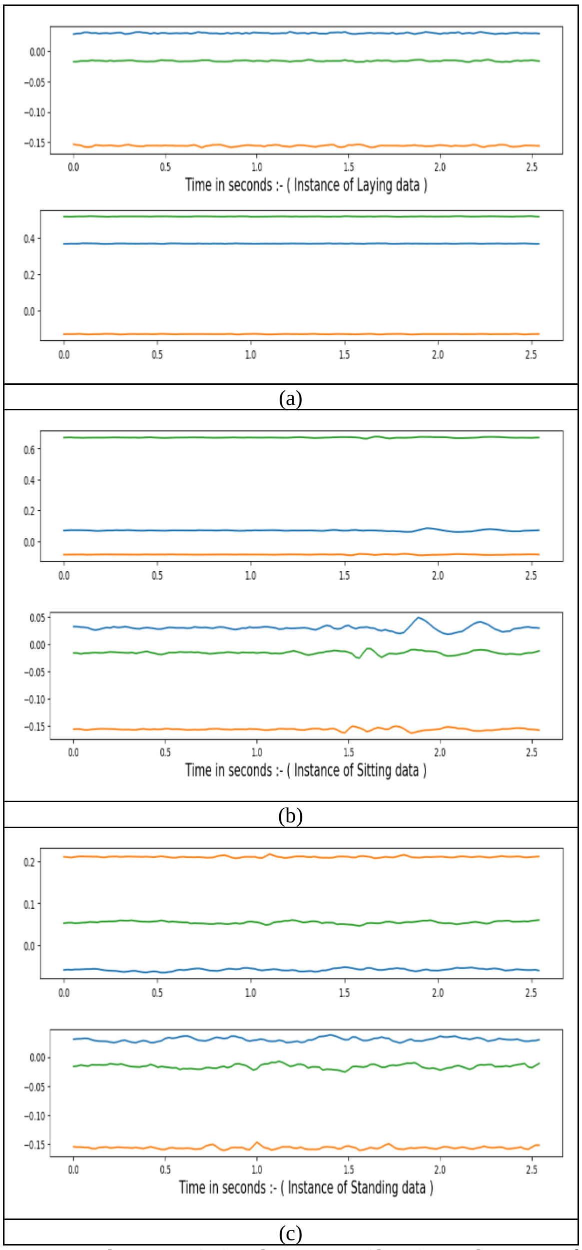

Fig. 8. Accelerometer (up) and gyroscope (down) signal patterns of a (a) laying sample, (b) sitting sample and (c) standing sample.

V. CONCLUSION

This paper proposes an efficient smartphone-based HAR model that leverages the collective competencies of multiple smartphone-embedded inertial sensors for a more inclusive feature analysis of elderly activities. The proposed deep learning framework is coined multi-stream TC-LSTM. In this model, the acquired raw data is firstly segmented into fixed-time windows using a sliding window technique. This enables a systematic examination of consecutive patterns, easing the analysis process. The segmented training data is then used to train and build an optimal model for feature extraction. Unlike other models, our proposed model engages independent temporal convolutional analysis on each inertial data, known as a convolutional stream. Since the employment of different sensors, the proposed model performs multiple temporal convolutional streams, allowing independent processing of inertial data of diverse sensing modalities. This enhances the model's ability to characterise the multisensory inertial data for more comprehensive feature extraction of diverse information. The extracted temporal features are further processed by LSTM for temporal abstraction. The results demonstrate that the proposed multi-stream TC-LSTM can obtain a promising performance in human activity recognition with an F1 score of 98.3%.

ACKNOWLEDGEMENTS

We acknowledge the Ministry of Higher Education (MOHE) for funding under the Fundamental Research Grant Scheme (FRGS/1/2020/ICT02/MMU/02/7).

REFERENCES

[1] V. Soni, H. Yadav, V. B. Semwal, B. Roy, D. K. Choubey and D. K. Mallick, “A Novel Smartphone-Based Human Activity Recognition Using Deep Learning in Health care,” *Lect. Notes Electr. Eng.*, vol. 946, pp. 493–503, 2023.

[2] O. Pavliuk, M. Mishchuk and C. Strauss, “Transfer Learning Approach for Human Activity Recognition Based on Continuous Wavelet Transform,” *Algorithms*, vol. 16(2), no. 77, 2023.

[3] S. Mekruksavanich and A. Jitpattanakul, “CNN-Based Deep Learning Network for Human Activity Recognition During Physical Exercise from Accelerometer and Photoplethysmographic Sensors,” *Lect. Notes Data Eng. Commun. Technol.*, vol. 117, pp. 531–542, 2022.

[4] G. Sebestyen, I. Stoica and A. Hangan, “Human activity recognition and monitoring for elderly people,” in *Proc. - 2016 IEEE 12th Int. Conf. Intell. Comput. Commun. Process.*, Cluj-Napoca, Romania, pp. 341–347, 2016.

[5] H. Cho and S. M. Yoon, “Divide and Conquer-Based 1D CNN Human Activity Recognition Using Test Data Sharpening,” *Sensors 2018*, vol. 18, no. 4, pp. 1055, 2018.

[6] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee and H. Moon, “Sensor-based and Vision-based Human Activity Recognition: A Comprehensive Survey,” *Pattern Recognit.*, vol. 108, pp. 107561, 2020.

[7] R. A. Voicu, C. Dobre, L. Bajenaru and R. I. Ciobanu, “Human Physical Activity Recognition using Smartphone Sensors,” *Sensors*, vol. 19, no. 3, pp. 458, 2019.

[8] D. Gholamiangonabadi, N. Kiselov and K. Grolinger, “Deep Neural Networks for Human Activity Recognition with Wearable Sensors: Leave-One-Subject-Out Cross-Validation for Model Selection,” *IEEE Access*, vol. 8, pp. 133982–133994, 2020.

[9] F. Hernández, L. F. Suárez, J. Villamizar and M. Altuve,

"Human Activity Recognition on Smartphones Using A Bidirectional LSTM Network," in *2019 22nd Symp. Image, Signal Process. Artif. Vision,* Bucaramanga, Colombia, pp. 1–5, 2019.

[10]C. Kok, V. Jahmunah, S. L. Oh, X. Zhou, R. Gururajan, X. Tao, K. H. Cheong, R. Gururajan, F. Molinari and U. R. Acharya, "Automated Prediction of Sepsis using Temporal Convolutional Network," *Comput. Biol. Med.*, vol. 127(C), pp. 103957, 2020.

[11]N. Nair, C. Thomas and D. B. Jayagopi, "Human Activity Recognition Using Temporal Convolutional Network," in *Proc. 5th Int. Workshop on Sensor-based Activ. Recogn. and Interact.*, no. 17, pp. 1-8, 2018.

[12]Y. W. Kim, K. L. Joa, H. Y. Jeong and S. Lee, "Wearable IMU-Based Human Activity Recognition Algorithm for Clinical Balance Assessment using 1D-CNN and GRU Ensemble Model," *Sensors 2021*, vol. 21, no. 22, pp. 7628, 2021.

[13]K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu and Y. Liu, "Deep Learning for Sensor-based Human Activity Recognition," *ACM Comput. Surv.*, vol. 54(4), no. 77, pp. 1-40, 2021.

[14]A. Michele, G. Biagetti, P. Crippa, L. Falaschetti, and C. Turchetti, "Recurrent Neural Network for Human Activity Recognition in Embedded Systems using PPG and Accelerometer Data." *Electronics*, vol. 10, no. 14, pp. 1715, 2021.

[15]S. Jameer and H. Syed, "Deep SE-BiLSTM with IFPOA Fine-Tuning for Human Activity Recognition using Mobile and Wearable Sensors," *Sensors 2023,* vol. 23, no. 9, pp. 4319, 2023.

[16]S. Raja Sekaran, P. Y. Han and O. S. Yin, "Smartphone-based Human Activity Recognition Using Lightweight Multiheaded Temporal Convolutional Network," *Expert Syst. Appl.*, vol. 227(C), pp. 120132, 2023.

[17]L. Trinh and B. Ha, "An Incorporation of Deep Temporal Convolutional Networks with Hidden Markov Models Post-processing for Sensor-based Human Activity Recognition," in *Proc. 11th Int. Sym. Inform. and Commun. Technol.*, pp. 96–102, 2022.

[18]Y. He and J. Zhao, "Temporal Convolutional Networks for Anomaly Detection in Time Series," *J. Phys. Conf. Ser.*, vol. 1213, no. 4, pp. 042050, 2019.