Sensor applications

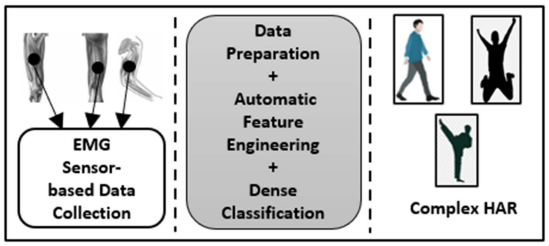
Enhanced Complex Human Activity Recognition System: A Proficient Deep Learning Framework Exploiting Physiological Sensors and Feature Learning

Nurul Amin Choudhury* and Badal Soni**

Department of Computer Science and Engineering, National Institute of Technology Silchar, Cachar 788010, India
**Graduate Student Member, IEEE*
***Senior Member, IEEE*

Manuscript received 3 October 2023; accepted 15 October 2023. Date of publication 19 October 2023; date of current version 30 October 2023.

Abstract—Human activity recognition is the process of identifying daily living activities of a person using sensor attributes and intelligent learning algorithms. Identifying complex human activities is tedious, as capturing long-term dependencies and extracting efficient features from the raw sensor data is challenging. This letter proposes an efficient and lightweight hybrid deep learning model for recognizing complex human activities using physiological electromyography (EMG) sensors and enhanced feature learning. The proposed convolutional neural networks - long short-term memory (CNN-LSTM) incorporates multiple 1-D convolution layers for spatial feature extraction and then feeds the generated feature maps to the recurrent layers to identify long-term temporal dependencies. Incorporating a physiological sensor-based raw EMG dataset and minimal preprocessing, we trained and tested our proposed model and achieved the highest accuracy of 84.12% and an average accuracy of 83%. The proposed model outperformed the benchmark models with optimal performance margins and generalized the patterns in significantly less computational time than other deep learning models.



Index Terms—Sensor applications, complex human activity recognition (HAR), deep learning, machine learning and daily living activities, physiological sensors.

I. INTRODUCTION

Human activity recognition (HAR) has witnessed significant improvements in recent years due to the development of wearable sensors and the advancements in smart learning approaches. HAR is the process of identifying an individual's daily living activities (DLA) using a set of sensors and advanced learning algorithms. Human activities are classified into normal and complex human activities based on muscle movements and energy consumption [1], [2]. DLA that has typical body or muscle movements and takes less energy to exert is termed normal DLA, and the activities that require frequent body movements and consume high energies are termed complex human activities [1].

Deep and machine learning models are widely adopted for sensor-based HAR due to their optimal performance and robust pattern recognition capabilities. Deep learning, in particular, has demonstrated superior performance over traditional machine learning methods, excelling in parameters, such as feature handling and more [1], [3]. Physiological sensors, such as electromyography (EMG), are increasingly employed in sensor-based HAR systems. EMG sensors capture the raw signals muscles generate during movement, providing valuable insights into an individual's electrical activity [4]. Unlike traditional wearable sensors that focus on external motions, EMG sensors can analyze muscle movement patterns and assertiveness, enabling more accurate recognition of complex activities. The strategic placement of EMG sensors on specific muscle groups allows for targeted analysis of activity-related muscle activations. As a result, EMG sensors are particularly valuable in applications such as sports performance analysis, physical rehabilitation, and gesture-based human–computer interaction.

Researchers around the globe practice HAR in multiple ways and segregate them on the basis of machine and deep learning algorithms. Coelho et al. [5] proposed a sensor-based deep and machine learning approach comprising shallow and convolution neural networks. Their model segregates various activities based on static and dynamic characteristics and classifies them using decision trees with manual feature extraction and convolutional neural network (CNN) with automatic feature handling, respectively. Mekruksavanich et al. [4] proposed a novel HAR system with knee abnormalities using surface EMG sensors and a goniometer. Al-qaness et al. [6] proposed a multilevel residual neural network for HAR on wearable sensors with an attention mechanism. Publicly available wearable sensor-based complex HAR datasets were considered for benchmark training and testing of the proposed model.

Koo et al. [7] developed a two-stream CNN feature extractor to detect human activities effectively by extracting the accelerometer and gyroscope feature separately for future concatenation. Normalized and concatenated attributes are then fed to the dense classification layer for activity classification using Softmax activation class probabilities. Thakur et al. [8] designed and implemented an autoencoder-based deep learning model for automatic feature engineering and HAR. Their model extracts local spatial features using CNN layers and comprises them into optimal feature subsets using autoencoders. They also

Corresponding author: Nurul Amin Choudhury (e-mail: nurul0400@gmail.com).
(Nurul Amin Choudhury and Badal Soni contributed equally to this work.)
Associate Editor: Chia-Chan Chang.
Digital Object Identifier 10.1109/LSENS.2023.3326126

incorporated a memory function for handling long-term dependencies and classified human activities using dense classification.

Most of the works on HAR are done on simple or standard human activities, and researchers around the globe managed to achieve efficient performance with their proposed methodology. Researchers [3], [9] worldwide exploit the development of HAR systems in multiple ways and mainly incorporate wearable sensors for collecting raw sensor data. However, the reliability and effectiveness of state-of-the-art methodologies cannot be inferred in the medical domain for patient rehabilitation and sports activity recognition for complex DLA. This letter proposes an efficient hybrid deep learning convolutional neural networks - long short-term memory (CNN-LSTM) model for classifying complex human activities using EMG physiological sensors.

The main novelties and contributions of this article are as follows.

1) A novel automatic feature engineering pipeline using CNN-LSTM has been designed and implemented for classifying complex human activities on raw EMG data without using intense data preprocessing pipelines and fine tuning.
2) A state-of-the-art EMG sensor-based physical action dataset is processed and exploited for recognizing complex human activities with normal activity instances using 8-channel electrical sensor information.
3) Multiple ensemble and deep learning approaches have been incorporated and exploited for detailed analytical discussion and benchmark comparison.

II. PROPOSED METHODOLOGY

The proposed methodology consists of multiple phases from data preparation to complex activity recognition using CNN-LSTM as follows.

A. Data Acquisition and Preparation

Multiple physiological sensor-based HAR datasets are available online for DLA recognition but do not have complex human activity instances. UCI EMG-based physical action dataset (UCI-PDA) [10] is one such dataset comprising 20 EMG-based complex and standard activity labels, collected from four individuals. UCI-PDA consists of the raw sensor data of three male and one female subject who experienced aggression, performing ten regular and ten aggressive activities in 20 individual experiments. The age of the considered subjects for the data acquisition is defined as Age $(A) = \{A = \mathbb{R}^+ | A > 24 \& A < 30\}$. Ethical constraints and safety precautions were followed on the basis of British Psychological Society to minimize the risks involved while performing the human activities.

Data were collected using a kickboxing standing bag and EMG apparatus in an experimental simulated environment. Eight electrodes were placed on the subjects' upper arms and legs, as shown in Fig. 1. The dataset consisted of 10 000 instances with eight attributes for each activity, representing body segments and corresponding muscles. There were 20 classes, including normal actions, such as *Bowing, Clapping, Handshaking, Hugging, Jumping, Running, Seating, Standing, Walking, Waving,* and aggressive actions, such as *Elbowing, Front kicking, Hammering, Headering, Kneeing, Pulling, Punching, Pushing, Side-kicking, and Slapping*. While the abovementioned data acquisition settings provide a controlled environment for data collection, the user perform activities are uncontrolled in nature as the subjects were allowed us to perform the DLAs as per their body types and capability. The incorporation of kickboxing was for the aggressive DLAs as the other day-to-day living activities were done without using the kick-boxing bag. However, due to the small user and age group, the dataset has limitations in terms of representatives and bias toward age and gender.

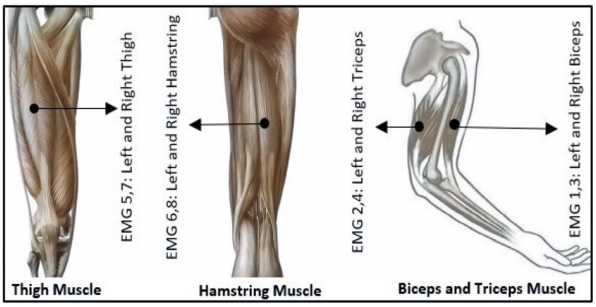


Fig. 1. Sensor mounting location for EMG-based data acquisition.

Furthermore, toward the data preparation, we minimally preprocessed the chosen data to make it classification ready for the proposed and benchmark models. We combined all the user's data and labeled it according to the utilization of the datasets' annotated metadata. After combining, we checked for missing and duplicate instances in the dataset and inferred that there were no missing values present in the dataset and that the dataset had a negligible amount of redundant instances. The need for feature analysis is not needed as our proposed CNN-LSTM model will automatically extract and select the spatial and temporal information for efficient classification. Also, upon exploring the dataset keenly, the class imbalance problem was not present in the synthesized dataset.

B. Proposed Model

The proposed model integrates CNN and LSTM layers for efficient feature engineering and classification using raw input data. The proposed model fetches the data from the input layer and passes them into the time-distributed convolution layer (16 filters, kernel size 2), followed by another CNN layer with the same parameters to extract local spatial features. A dropout layer prevents overfitting, and max-pooling reduces feature dimensionality as mentioned (1)–(3), respectively. The flattened spatial features are then fed into a recurrent network with memory and neuron units for capturing temporal data

$$X = \{x_1, x_2, x_3, \ldots, x_n\} \mid \; x_i \in \mathbb{R}^d \tag{1}$$

$$C = \sigma(W * X + b) \mid * \rightarrow 1D \text{ Conv. Operation} \quad W \rightarrow \text{Filter Weights}, \; b \rightarrow \text{Bias} \tag{2}$$

$$F(C) \text{ as } C = \{c_1, c_2, \ldots, c_F\} \mid F \rightarrow \; 1D \text{ MaxPooling}. \tag{3}$$

The LSTM layers capture temporal dependencies and long-term patterns for generalizing EMG patterns [9], [11]. We included additional dropout and dense layers with recified linear unit (ReLu) activation for feature selection. A final dropout layer and softmax classifier were added to produce recognition class probabilities. The CNN-LSTM model was trained using categorical cross-entropy loss and the Adam optimizer. Fig. 2(a) illustrates the model's architecture.

Furthermore, the proposed methodology does not incorporate fine-tuning and intense data preprocessing, as most pretrained models are tailored to the specific data domain. Also, the incorporated time-series data displays unique spatial and temporal dependencies that necessitate significant model modifications. Additionally, intensive data preprocessing introduces computational overhead during training and real-time testing due to the need for identical preprocessing steps, potentially causing delays in recognition results.

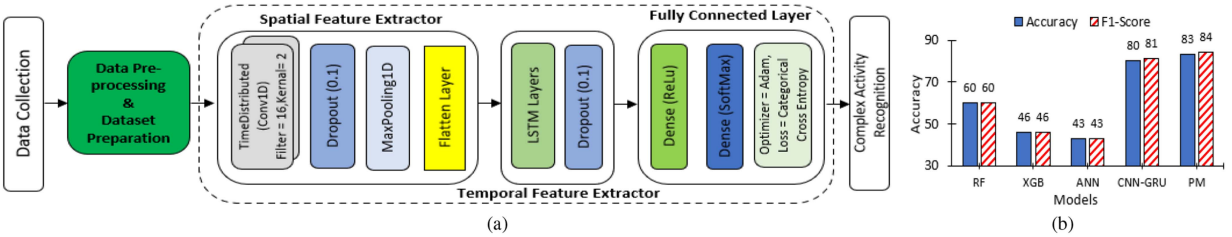


Fig. 2. Architecture of the proposed model with performance comparison graph. (a) Architecture of our proposed hybrid CNN-LSTM model. (b) Accuracy and F1-score comparison of proposed model with benchmark models.

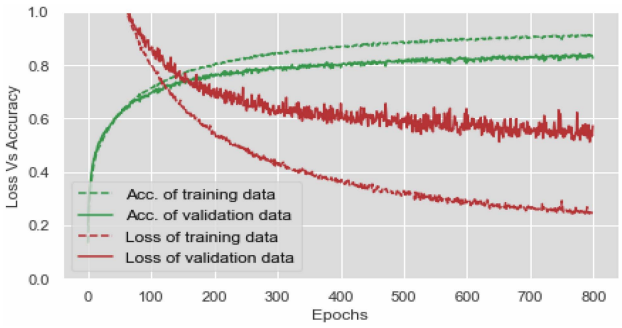


Fig. 3. Training and validation accuracy versus loss plot of proposed CNN-LSTM model.

C. Evaluation Metrics

To test and benchmark our proposed model, we incorporated multiple ensemble and deep learning models for performance comparison on various evaluation metrics, such as precision (P), recall (R), F1-score (F1), accuracy (A), computational time (C_T), and performance loss. Along with the performance measure, we incorporated multiple machine [random forest (RF) and extreme gradient boosting (XGB)] and deep learning (artificial neural network (ANN), convolutional neural networks - gated recurrent unit (CNN-GRU)) models for benchmark testing and comparison. These models were incorporated to comprehensively evaluate different learning paradigms' performance, including traditional ensemble and deep learning techniques prominently exploited for human activity recognition systems.

III. EXPERIMENTAL RESULT AND DISCUSSION

The proposed model was developed in Python using Jupyter Notebook as the primary editor of choice. To train, validate, and test the developed model, we used a Fujitsu workstation consisting of a 3.6 GHz Xeon-W 2133 multicore processor, 32 GB RAM, and 2 TB HDD. GPU was not used for the model training as we have sufficient computational power with our incorporated processor.

A. Performance Analysis of Proposed Model

After formulating a complex activity dataset and performing data preprocessing, we trained and validated our proposed CNN-LSTM model for activity classification. Upon testing, the proposed model achieved the highest accuracy of 84.12% and average accuracy of 83%. The loss on testing data was also very minimal from the benchmark models, with an average loss of 0.515, as illustrated in Fig. 3. Benchmark modes, such as RF, XGB, ANN, and CNN-GRU achieved lower performance accuracies than our proposed model with average accuracies of 58%, 45%, 42%, and 80%, respectively.

The accuracy comparison of all the benchmark models with our proposed CNN-LSTM model is illustrated in Fig. 2(b). The combination of convolution and memorization scheme allows our model to recognize complex patterns in human activity data, leading to improved accuracy and performance compared to traditional machine learning models that struggle to capture intricate spatial and temporal relationships of raw sensor data. Deep learning models, such as ANN and CNN-GRU, suffer from model overfitting and cannot generalize the data patterns optimally. Likewise, RF and XGB failed to memorize the training data efficiently and performed poorly during model testing, as described in Table 1.

B. Loss and Computational Time Analysis

The proposed model achieved optimized computational time during the training and validation of raw sensor instances by incorporating small filter sizes and a reduced number of filters in the convolution phase. After spatial feature extraction, regularization modules, such as Dropout and MaxPooling further optimized the training process by reducing feature map size. Additionally, Dropout and rectified linear unit throughout the feature engineering layers enabled the model to generalize nonlinear relationships without increasing complexity. As a result, the proposed model achieved an efficient computational time of 4412 ± 4434 fractional seconds (Fs), outperforming other deep learning approaches, such as ANN and CNN-GRU, which recorded significantly higher computational times of 6621 ± 6650 and 7431 ± 7521 Fs, respectively.

As described in Table 1, benchmark models, such as ANN and CNN-GRU, faced comparatively high training loss and lower precision rates. The lack of features and their intuition in machine learning models (RF and XGB) forced the model to recognize complex DLA using their probabilistic function, leading to poor classification. ANN fails to capture spatial (ex—orientation, rotational rate, maximum acceleration, etc.) and temporal (ex—mean acceleration over time, peak orientation over time, etc.) features as it lacks convolution and memory function for feature engineering and loses most of the performance while validating the training set. Furthermore, with CNN-GRU, it acquired the time and frequency domain features but failed to generalize the data patterns optimally as it overfits itself during model training. The need for huge training data instances, which is very difficult to collect and process makes the model's generalizability poor and it overfits while generalizing on small-length training sequences.

C. Ablation Study

The ablation study showcased the essential roles of the critical components incorporated in the proposed model, as described in Table 2. Removing the LSTM layers significantly decreased the accuracy to 74%, highlighting the crucial role in capturing temporal dependencies using the memorization function. While removing dropout layers,

TABLE 1. Detailed Performance Comparison of Our Proposed Model With Incorporated Benchmark Models

Classifier/metrics	Highest Accuracy (%)	5-Fold accuracy.(%)	Avg. metrics	P	R	F1	Test loss	Ct (Frac. Sec.)
RF	60.38	58	Micro Avg.	0.60	0.60	0.60	321 ± 353	280 ± 308
			Weighted Avg.	0.60	0.60	0.60		
XGB	46.21	45	Micro Avg.	0.46	0.47	0.46	397± 412	3600 ± 3635
			Weighted Avg.	0.46	0.47	0.46		
ANN	43.41	42	Micro Avg.	0.44	0.43	0.43	420 ± 436	6621 ± 6650
			Weighted Avg.	0.44	0.43	0.43		
CNN-GRU	81.02	80	Micro Avg.	0.81	0.81	0.81	0.510 ± 0.548	7434 ± 7521
			Weighted Avg.	0.81	0.81	0.81		
Prop. CNN-LSTM	84.12	83	Micro Avg.	0.84	0.85	0.84	0.505 ± 0.538	4412 ± 4434
			Weighted Avg.	0.84	0.85	0.84		

TABLE 2. Ablation Study of the Proposed CNN-LSTM Model

Model's ablation	5-fold accuracy (%)	C_t (Frac. Sec.)
Prop. model without LSTM	74	3257 ± 3420
Prop. model without Dropout	78	7671 ± 8021
Prop. model without CNN	80	8561 ± 8666
Prop. model	83	4412 ± 4434

the model generalization evidenced a slight accuracy drop to 78% as the neuron starts to overfit the model generalization. One of the key observations was the drop in accuracy to 80% when CNN layers were removed, underlining their importance in extracting spatial features from the input data. These findings highlight the critical balance between temporal and spatial modeling elements, emphasizing the need for thoughtful integration of LSTM and CNN layers to enhance the model's performance and optimization.

IV. MODEL'S GENERALIZABILITY AND LIMITATIONS

The proposed CNN-LSTM model demonstrated strong generalizability and outperformed several benchmark models in terms of accuracy and computational efficiency. The performance of the proposed model indicates the ability to recognize complex patterns in human activity data effectively. Unlike traditional learning models—RF and XGB, which struggled to memorize training data efficiently, the CNN-LSTM model excelled in learning and recognizing complex patterns. Moreover, the model's efficient computational time of 4412 ± 4434 Fs during training demonstrated its practical feasibility. This efficiency outperformed other deep learning models, such as ANN and CNN-GRU, which required significantly more computational time.

Despite the model's strength in performance and computational times from the benchmarks, the model's performance depends on the quality and quantity of training data instances. Also, while the model's computational efficiency is advantageous, it may not be suitable for deployment on resource-constrained devices or in real-time applications that demand extremely low latency.

V. CONCLUSION

This letter proposes an efficient and streamlined hybrid CNN-LSTM model using EMG sensors for complex human activity recognition. A state-of-the-art physiological sensor-based HAR dataset was formulated with 20 complex human activities and 8-channel EMG sensors. Our proposed model achieved an average accuracy of 84.12% and the highest accuracy of 83%. The validation and test set loss was also minimal compared to incorporated benchmark models. Furthermore, the proposed model optimizes the computational overhead and trains the network structure in significantly less computational time than other benchmark deep learning models, such as ANN and CNN-GRU.

In future, we will develop a lightweight deep learning-based feature fusion pipeline for complex HAR in optimized training iterations and computational time.

REFERENCES

[1] N. A. Choudhury and B. Soni, "In-depth analysis of design & development for sensor-based human activity recognition system," *Multimedia Tools Appl.*, pp. 1–40, 2023.

[2] E. Ramanujam, T. Perumal, and S. Padmavathi, "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13029–13040, Jun. 2021.

[3] N. A. Choudhury and B. Soni, "An adaptive batch size based-CNN-LSTM framework for human activity recognition in uncontrolled environment," *IEEE Trans. Ind. Informat.*, vol. 19, no. 10, pp. 10379–10387, Oct. 2023.

[4] S. Mekruksavanich, P. Jantawong, N. Hnoohom, and A. Jitpattanakul, "Human activity recognition for people with knee abnormality using surface electromyography and knee angle sensors," in *Proc. Joint Int. Conf. Digit. Arts, Media Technol. ECTI Northern Sect. Conf. Elect. Electron. Comput. Telecommun. Eng.*, 2023, pp. 483–487.

[5] Y. L. Coelho, F. D. A. S. D. Santos, A. Frizera-Neto, and T. F. Bastos-Filho, "A lightweight framework for human activity recognition on wearable devices," *IEEE Sensors J.*, vol. 21, no. 21, pp. 24471–24481, Nov. 2021.

[6] 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.

[7] I. Koo, Y. Park, M. Jeong, and C. Kim, "Contrastive accelerometer-gyroscope embedding model for human activity recognition," *IEEE Sensors J.*, vol. 23, no. 1, pp. 506–513, Jan. 2023.

[8] D. Thakur, S. Biswas, E. S. L. Ho, and S. Chattopadhyay, "Convae-LSTM: Convolutional autoencoder long short-term memory network for smartphone-based human activity recognition," *IEEE Access*, vol. 10, pp. 4137–4156, 2022.

[9] N. A. Choudhury and B. Soni, "An efficient and lightweight deep learning model for human activity recognition on raw sensor data in uncontrolled environment," *IEEE Sensors J.*, vol. 23, no. 20, pp. 25579–25586, Oct. 2023.

[10] T. Theodoridis, "EMG physical action data set," *UCI Mach. Learn. Repository*, 2011, doi: 10.24432/C53W49.

[11] N. A. Choudhury and B. Soni, "An efficient CNN-LSTM approach for smartphone sensor-based human activity recognition system," in *Proc. 5th Int. Conf. Comput. Intell. Netw.*, 2022, pp. 01–06.