Proc. of International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA 2024)
1-2 February 2024, Victoria-Seychelles

Analysis of Machine Learning-based Human Activity Prediction Model for Assistive Exoskeleton

Shubhashisa Sahoo
Centre for Artificial Intelligence and Robotics
Bangalore, India
sahoo@ieee.org

Mayank Sharma
Renaissance Technology Private Limited
New Delhi, India
mayanks.in@ieee.org

Aniket Yadish
Dept. of Mechanical Engineering PES University
Banglore, India
yadish.aniket@gmail.com

Sawan Kumar
Renaissance Technology Private Limited
New Delhi, India
nawaskumar@gmail.com

***Abstract*—Musculoskeletal disorders significantly affect the general population, especially those involving lower back discomfort. This is the most important reason for the decline in work efficiencies in industrial and defense applications. Under these conditions, an active exoskeleton will be quite helpful. Selecting an appropriate control strategy for an active exoskeleton requires the prediction of human motion profiles. However, the complex and dynamic nature of the exoskeleton model limits the model-based approach. Thus, data-driven parameter estimation models are preferred in this scenario. Human Activity Recognition (HAR) research has recently focused on deep neural networks. There is a noticeable delay between the observed motion profile and detected human activity due to the intrinsic latency in the exoskeleton control architecture. As a result, HAR is not enough to achieve the objective on its own. On the contrary, there is not much prior study on estimating and predicting human motion profiles. This study presented a brief comparison of the widely used predictive models. The authors emphasized human activity profile prediction by employing deep learning techniques to compensate for the lag and facilitate the real-time control strategy. According to their performance comparison, the Long Short-Term Memory (LSTM) model performed better than other neural networks and provided the highest prediction accuracy.**

Keywords*—Exoskeletons, Human Activity Recognition, Deep Learning, Machine Learning

I. Introduction

Exoskeletons are wearable structures designed to support human movements and increase an individual's strength and capabilities [1]. The exoskeleton systems can be categorized into rehabilitation, assistive, and augmentation exoskeletons. The rehabilitation exoskeletons are meant to automate a series of rehabilitation exercises to keep track of the improvement of patients. The assistive exoskeletons come in between the rehabilitation and augmentation exoskeletons. This exoskeleton mainly serves operational modes: passive support, which keeps the limb in a set position for a designated time; active mode, where it reacts to the user's intentions for specific movements or trajectory completion during tasks; and the ability to steer or compel arm placement on a predefined path. Augmentation exoskeletons are designed and engineered to enhance human performance in demanding industrial environments. In this scenario, the user operates the device, and the exoskeleton bears the entire weight of the manipulated objects. This study is restricted to the assistive exoskeletons, where the powered actuators share the majority of the load [2]. Human activity recognition (HAR) has a substantial role in active exoskeletons by providing physical activity prediction using dedicated sensors.

Predicting human motion is of paramount importance in assistive exoskeleton design. This precision in forecasting human movement enables the exoskeleton to provide proactive and smooth assistance to the user, leading to a more instinctive and effortless interaction. Furthermore, exoskeletons that can anticipate and adjust to the user's motion alleviate the cognitive load on the wearer by eliminating the need for constant device adjustments. This ability to foresee the user's movements also prevents unintended motions and collisions, enhancing safety. The studies on HAR in the context of the powered exoskeleton are concerned with the estimation of the human pose and gait phases based on various motion profiles [3]. These profiles are initially recorded with dedicated sensors like Inertial Measurement Units (IMU), Force-Sensitive Resistors (FSR), surface Electromyography (sEMG), and pressure pads. Based on these recorded signals, the HAR model can estimate the activity. The estimated activity can be utilized as a reference signal to the control unit of the exoskeleton actuators.

An inherent latency is imposed primarily due to computation time in the HAR model and the control unit. Consequently, the generated control signal will always lag compared to the real-time actual reference. Therefore, a deep learning-based motion profile estimation model is proposed to predict the motion profile in advance to compensate for the lag between the actual and estimated signal. The contributions of this study are as follows.

- Carried out a thorough analysis of the body of research on human activity recognition to pinpoint the most recent developments, trends, and problems.
- Meticulously chosen and pre-processed pertinent open-source HAR dataset covering a range of human activities, guaranteeing a representative sample for thorough algorithm assessment.
- Implemented and fine-tuned a variety of deep learning algorithms commonly used in HAR.

- The performance of the various models is compared based on matrices like R2 score, Mean Squared Error (MSE), and Mean Absolute Error (MAE). This comparison helps to select the best-suited model for the problem at hand.

The organization of the paper is as follows. Section II describes the existing literature survey. Section III includes the description of data samples and their techniques. The various deep learning models considered for the study are included in section II. The results and discussions are explained in section IV, and section V concludes the studies.

II. Existing Studies

Several prior research discussed the usage of HAR in exoskeletons. The study in literature [4] emphasized developing a payload estimation model that classifies weights of 5kg, 10kg, and 15kg. The classifier was implemented on 12 volunteers. The HAR model reached an F1 score of 89.49% for the action prediction, which includes activities stand, walk, and payload interaction, and 96.33% for lifting and lowering activities, respectively. However, this study is only limited to the HAR purpose, which may not be efficient in real-time scenarios. The authors in the literature [5] have discussed body sensor-based activity modeling and recognition. Twelve activities like walking, lifting, and running were performed.

Long short-term memory (LSTM)-based neural structured learning (NSL) model was used with kernel discriminant analysis (KDA) for enhancing the features. The proposed approach by the authors achieved around 99 % recall. In the literature [6], the authors have validated the feasibility of real-time HAR with the wearable exoskeleton robot [7]. Considering the eight activities, five deep-learning models have been trained and used for HAR. The overall recognition accuracy was 97.35%, with an inference time under 10ms. The study in literature [8] is based on HAR using radars. Various deep-learning approaches were used for activity recognition in employing radar signals and found that the recognition accuracy was within the acceptable range. In the literature [9], the authors have proposed a two-stage genetic algorithm-based feature selection algorithm with a fixed activation number (GFSFAN). Time series data from 9 daily activities are collected for modeling. While comparing with other feature selection algorithms, the result prevailed that the proposed algorithm can discard redundant features and maintain comparable classification performance.

In the literature [10], authors have used deep recurrent neural networks (DRNNs). The results show that the proposed method outperforms the conventional machine learning methods, such as support vector machine (SVM) and k-nearest neighbours (KNN). The literature [11] employed CCTV videos and camera images for detecting human activities. They utilized experimental data containing 5648 images from a camera. They also employed the Convolutional Neural Network (CNN) Classifier to recognize human activities with a very good accuracy of 99.82%. However, most of the studies are related to HAR. Only very few articles are available related to motion profile prediction.

In the literature [12], the proposed MoveNet neural network consists of auto-encoder-based architecture with an additional feature mapping architecture that connects the encoder and decoder section. The data set consists of three constant walking speeds of 0.8, 1, and 1.2 m/s and nine AOIs ranging from −10° to +10° at an increment of 2.5°, recorded on a Bertec instrumented split-belt treadmill. For each condition, the gait data are recorded for 60 sec at a sampling rate of 100 Hz by a ten-camera Vicon motion capture system.

Most of the literature, including the literature discussed in this section, is mainly focused on the human activity classification task, which doesn't anticipate the future human motion profiles that can have a crucial impact on the exoskeleton control design. Therefore, this study emphasizes motion kinematics forecasting using deep neural network models.

III. Methodology

This study is focused on the comparative analysis of the various deep learning models in the context of HAR. A brief explanation of the methodology is depicted in the Fig. 1.

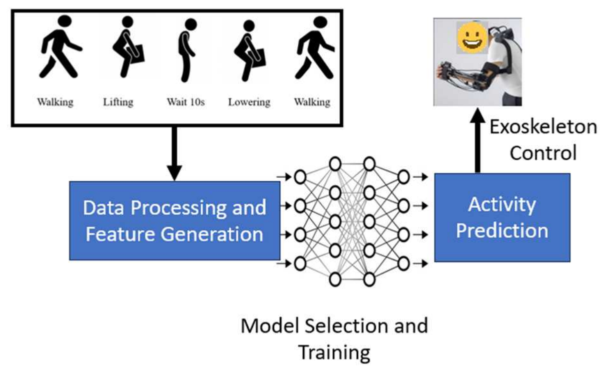

Fig. 1. Working principle of the proposed method.

As depicted in Fig. 1, the first stage of the study is the data curation and preprocessing. Then, the extracted features from the preprocessed data are used to train and evaluate the model. Then, the model's predicted human motion profile is taken as the reference profile in the exoskeleton actuator control model. The predicted motion profile is superior to the estimated motion profile because of the inherent latency in the motion estimation system in terms of hardware and software. Therefore, a carefully predicted future motion profile can compensate for this latency. This study is mainly focused on data preprocessing, model training, and evaluation that can be useful in designing the various exoskeleton controllers. The detailed explanation of data preprocessing, model selection, training, and evaluation are as follows.

A. Data Preparation

This study evaluates various state-of-the-art HAR models based on their prediction performance. An open-source data set is selected for an unbiased comparison of these models. The details of the data samples and the baseline HAR models are discussed as follows. An open-source dataset mentioned in the literature [4] is considered in this study. The dataset contains three linear acceleration signals and three rotational velocity signals from five IMUs placed on different locations of the subject. The considered activities are bending, walking, and lifting. Twelve different subjects, six male and six female, were considered to follow specific data curation protocol. Each subject walked, lifted a weight, waited for ten seconds, dropped the weight, and resumed walking. The data were sampled at 100 Hz, and offline data processing was done. In this study, the motion prediction in the sagittal plane is considered. Therefore, the data from the five gyroscopes contributing only to this plane is selected. The windowing algorithm is employed on the given dataset for sampling temporal features, as depicted in Fig. 2.

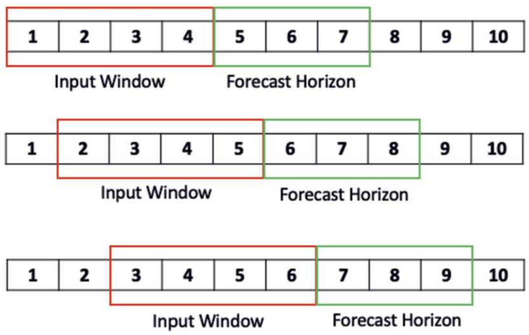

Fig. 2. Windowing scheme for dataset.

Fig. 3 depicts the input and output samples obtained by windowing gyroscopic data from the IMU sensor placed on the thighs. A total of 129614 samples are taken with an input window size of 100, and an output window size is 15. The input shape is considered as m×n. Here, m is the number of data samples, n is the number of features. The output shape is (s× out_{seq} x n) where s is given as,

$$s = \left(\frac{m-(inp_{seq}+out_{seq})}{stride} + 1\right) \tag{1}$$

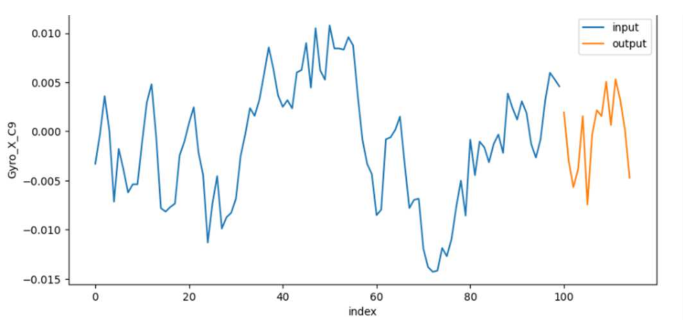

Fig. 3. Temporal samples of gyroscope.

In (1), inp_{seq} and out_{seq} are input and output sequences. Stride depends on the overlapping between the consecutive windowed samples. Hence, the input shape is (129040, 500), and the output shape is (129040, 75). Normalization was performed on the windowed data. Min-max normalization was used, which is given as,

$$x(scaled) = (x - x_{min}) / (x_{max} - x_{min}) \tag{2}$$

where, x_{min} and x_{max} are the minimum and maximum values of the features under consideration.

B. Model Consideration

Many researchers have shown keen interest in HAR. Prior studies leverage sensors, like gyroscopes and accelerometers, for human activity prediction. Despite the abundance of machine learning and deep learning models, only a handful are used due to their performance and popularity. This study has discussed some of the deep learning and machine learning models that are used for HAR. The baseline HAR models are reconfigured as per the requirement of predictive modeling in the context of this study. The baseline models based on existing studies are as follows:

1) K-Nearest Neighbours (KNN): This method is used for classification and prediction. Since it does not make any presumption on the elementary dataset, it is an instance-based algorithm [13]. *Generally following distances are used for calculating k nearest neighbours:*

$$\text{Euclidean distance } \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2} \tag{3}$$

$$\text{Manhattan distance } \sum_{i=0}^{n}|x_i - y_i| \tag{4}$$

2) Artificial Neural Network (ANN): It is inspired by biological neurons, which consist of interconnected nodes and are organized in layers. These layers are of three types input layer, hidden layer, and output layer [14] as shown in Fig. 4. The relation between output y_t and the inputs $x_{t-1}, x_{t-2} \ldots\ldots x_{t-p}$ is given by following the mathematical model [13].

$$y_t = g(\beta_{0j} + \sum_{i=1}^{p}\beta_{ij}x_{t-1}) \tag{5}$$

where, β_{ij} are the model parameters, also known as connection weights, p is the number of input nodes, and g is the hidden layer activation function.

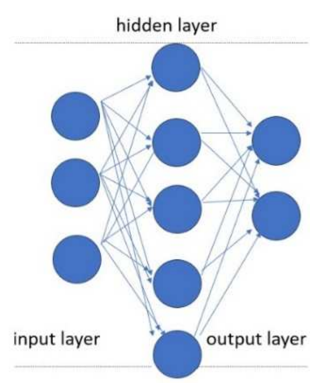

Fig. 4. Schematic diagram of ANN.

3) Recurrent Neural Network (RNN): It is a neural network where the output from the previous step is fed to the current time stamp computations, as shown in Fig. 5. Since it remembers the previous input to the network, the hidden state or memory state is the main feature of RNN.

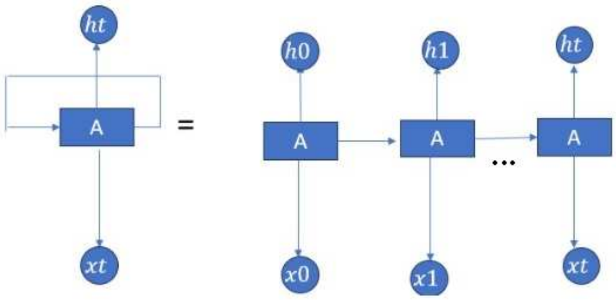

Fig. 5. The structure of RNN.

A typical RNN is shown in Fig 5. The i^{th} neuron can be described by the following equation [15].

$$x_i(k+1) = y_i + x_i(k) + \sum_{j=1}^{n_i+n_h} \tilde{w}_{ij} + z_j(k) \tag{6}$$

4) Long-Short-Term Memory (LSTM): It is a type of RNN that is used for handling sequential data. They are capable of handling long-term dependencies. Apart from hidden state h_t and an internal or cell state c_t, it also contains three gates; input i_t, forget f_t, and output gate o_t as shown in Fig. 6.

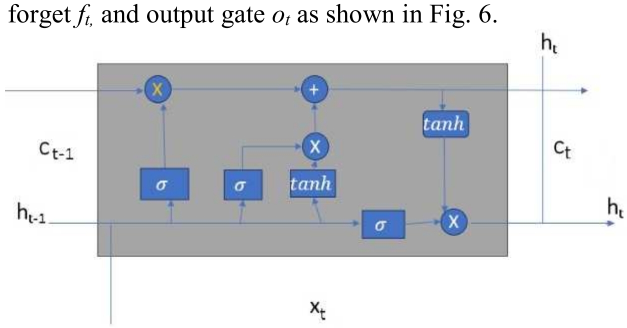

Fig. 6. An illustration of an LSTM.

The parameters x_t is the input at t^{th} time stamp and h_{t-1} is the previous hidden state. Following equations can be used to find out the current hidden state h_t [16].

$$f_t = sigmoid(W_f[h_{t-1};x_t]+b_f) \tag{7}$$

$$i_t = sigmoid(W_i[h_{t-1};x_t]+b_i) \tag{8}$$

$$o_t = sigmoid(W_o[h_{t-1};x_t]+b_o) \tag{9}$$

$$c_t = f_t \otimes c_{t-1} \oplus i_t \otimes \tanh(W_c[h_{t-1};x_t]+b_c) \tag{10}$$

$$h_t = o_t \otimes \tanh(c_t) \tag{11}$$

where, w_f, w_i, w_o, w_c, b_f, b_i, b_o, b_c represent the weights and biases, respectively. Employing (7) to (8), the hidden state for the current timestamp is given as follows.

$$h_t = f(h_{t-1}, x_t) \tag{12}$$

5) Convolutional Neural Network (CNN): It is another type of neural network which is primarily used for computer vision. It is a feedforward kind of neural network. Usually, the hidden layers consist of many convolutional layers, as shown in Fig. 7. They are extended versions of ANN [17].

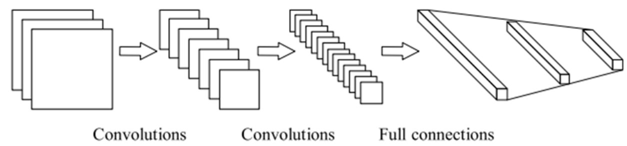

Fig. 7. Illustrates the architecture of CNN [17].

In this model, the dedicated filters are convolved with the input tensors to automatically extract the appropriate features that can be utilized further by a fully connected network for classification or regression problems.

IV. Result and Discussion

A. Results

This section discusses the performance of various ML and DL models. The details of the various models and their training parameters are discussed as follows. The KNN model variant, i.e., K-neighbours regressor is used considering 5-neighbours. After several trials, the optimum model is achieved using manual hyperparameter optimization. The optimum model is compared with the other baseline models.

The considered ANN model with one hidden layer is employed with linear activation output. The adaptive moment estimation (ADAM) optimizer function is also used with 100 epochs. Fig. 8 shows the training loss and validation loss (On a scale of 1.80 - 3.00) for the ANN model. The figure shows both training loss and validation loss decrease at the same rate and approach each other after 100 epochs.

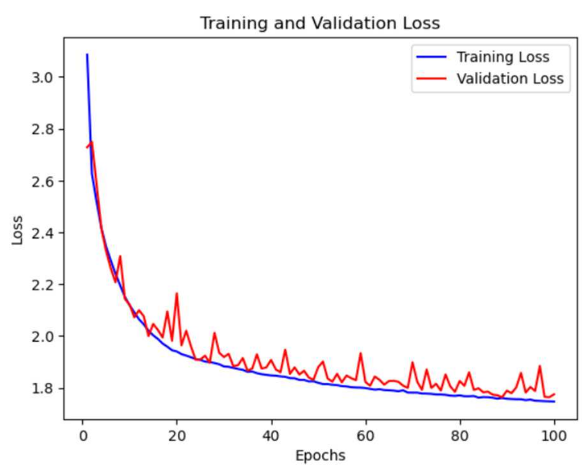

Fig. 8. Training and validation loss for ANN.

In the RNN model, an RNN layer connected with five dense layers is deployed each with 75 units and L2 regularization. The model uses ADAM optimizer and the model has been trained for 10 epochs with a batch size of 32. The RNN models' training and validation curve is presented in Fig. 9. The RNN model is trained with neurons in the hidden layers and a fully connected network connected at the last RNN cell, RMS Prop. The optimization algorithm is adopted with a learning rate (LR) of 0.005. The fully connected layer delivers the forecasted output. The learning curve of this model prevails that at the 100^{th} epoch, the training MSE and validation MSE are 2.5 and 0.5, respectively. After the 100^{th} epoch, the train and validation loss became saturated, and no further learning was observed. The LSTM model has an LSTM layer with 64 units/neuron. The output of the LSTM layer is connected to a dense layer with 75 neurons.

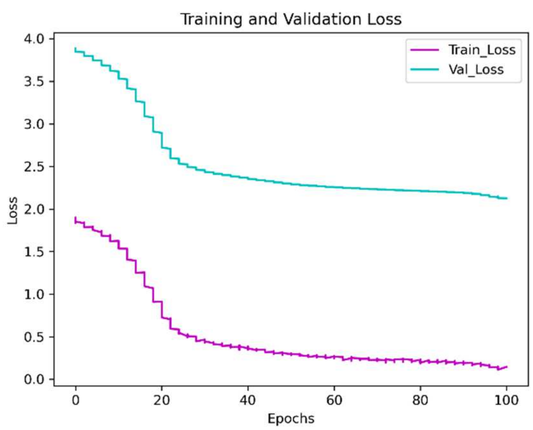

Fig. 9. Training and validation loss for RNN.

Furthermore, the model is compiled with ADAM optimizer (LR-0.01) and uses MSE as a loss function. The LSTM plot for training and validation loss is shown in Fig. 10. The result prevails that over the 100^{th} epoch, training loss, and validation loss are significantly low compared to the ANN and RNN models. At the end of the 100^{th} epoch, the train and validation losses are 0.08 and 0.04, respectively, which are very close to each other.

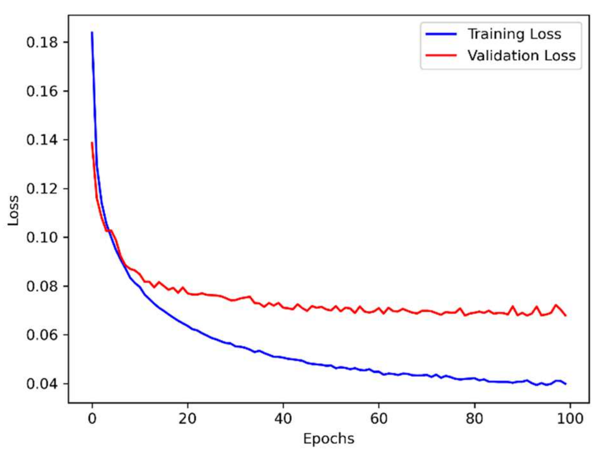

Fig. 10. Training and validation loss for LSTM.

Therefore, it is considered that the model is optimized for the given data samples. The CNN models are unconventional for the time series analysis; however, 1D CNN can achieve good prediction outcomes. Since several prior ats have advocated the usage of CNN for HAR, this model is also considered in this study. The CNN model with a 1-D convolutional layer with 32 filters, a kernel size of 3, and ReLU activation is used. A flattened layer, four dense layers with 64,32,32,32 neurons, respectively, employing the ReLU activation function, are considered. The model is trained with 100 epochs using the ADAM optimization technique.

The training and validation loss (on a scale of 1.30 - 1.80) curve for the CNN model is shown in Fig. 11. From the figure, it is evident that after the 100^{th} epoch, the reduction in training and validation sets are uneven, i.e., the training loss decreases significantly compared to validation loss. This signifies that the model the considered model is overfitting. Although various regularization and batch normalization techniques were evaluated, but no significant improvement was observed. It requires more training samples for further improvement in the prediction performance.

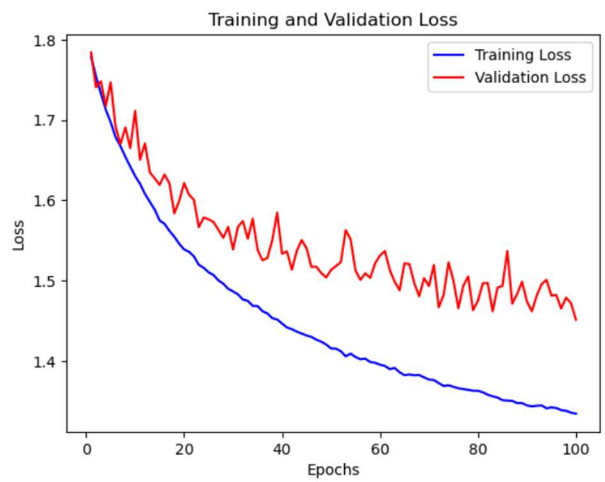

Fig. 11. Training and validation loss for CNN.

The forecasted result is depicted in Fig. 12. The comparison of the actual future samples and the predicted samples indicates a very small deflection with an R2 score of 0.90. Therefore, the LSTM is considered a suitable model for the objective of this study.

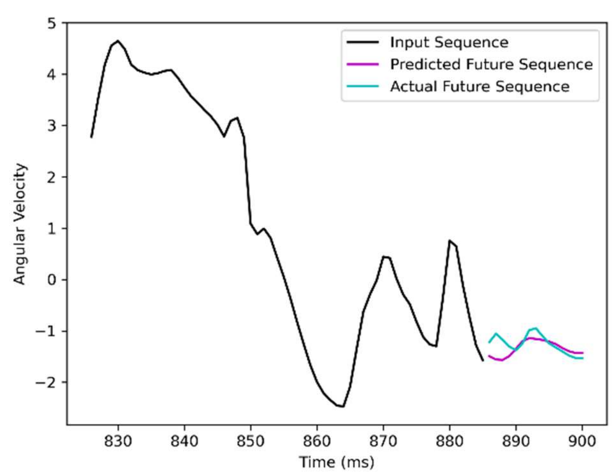

Fig. 12. LSTM forecasting evaluation.

B. Discussion

In this section, the training and validation performance of various common prediction models are compared. Multiple metric-based evaluation criteria are considered. The prediction results for the validation and training set are shown in Table I and Table II, respectively. Table I evaluates the performance of baseline models on the training set. This table shows that the LSTM has the best performance in terms of MSE, MAE, and R2 scores compared to the other models.

In this study, it is found that the LSTM model outperformed the other models under consideration in the context of training samples. In addition to having the lowest MSE and MAE scores which are critical for prediction performance, LSTM secured the highest R2 value. In terms of total training time, the model performance is not well compared to ANN and CNN, because the LSTM model architecture considered in this study is more complex and larger in size compared to the ANN and RNN.

TABLE I. BASELINE MODELS EVALUATION ON TRAINING DATA SET

Model used	R2_score	Mean squared error (MSE)	Mean absolute error (MAE)	Training time (s)
ANN	0.652	2.894	0.716	420
RNN	0.57	2.55	1.35	1111
CNN	0.643	2.67	3.45	220
LSTM	**0.895**	**0.05**	**0.12**	**840**

Table II describes the performance of the baseline models on the validation set. As observed from the table, LSTM gives the best results amongst other models in terms of R2 score, MSE, and MAE. However, the inference time of the LSTM model is highest among all i.e., 0.178 seconds; this can be compensated with the less prediction error advantage.

TABLE II. BASELINE MODELS EVALUATION ON VALIDATION SET

Model used	R2_score	Mean squared error (MSE)	Mean absolute error (MAE)	Inference time (s)
ANN	0.632	1.765	0.742	0.035
RNN	0.652	2.082	0.815	0.150
CNN	0.647	1.777	0.716	0.022
LSTM	**0.900**	**0.08**	**0.14**	**0.178**

The comparison of the best baseline model with the existing predictive model discussed in the literature review is presented in Table III. The comparison shows improvement in the prediction performance in terms of MSE and MAE for the baseline model compared to the existing predictive model.

TABLE III. COMPARISON OF BASELINE MODELS WITH PRIOR STATE OF THE ART

Model used	Mean squared error (MSE)	Mean absolute error (MAE)
MoveNet [12]	3.24	2.66
LSTM	**0.08**	**0.14**

It is evident from Tables II and III that LSTM produces more accurate results than other models. The model is found to be optimized for the bias and variance issues after 100 epochs, as demonstrated by the training loss and validation loss in Fig. 9. Although, the deep learning-based models like LSTM have shown good performance in terms of prediction accuracy in the context of HAR, these models faces certain limitations as follows.

- Due to limited data availability of human activity, only a few activities can be realised through these models.
- Limited data availability causes a biased model because less attention can be paid to the edge case.
- Larger feature space may require a large model that may increase the space and time complexity.

V. CONCLUSION

Industrial exoskeletons have great potential to reduce human effort, especially in the defense sector. Therefore, in this study, various deep learning models are evaluated for crucial activities in the context of weight lifting, walking with weight, lowering, and lifting. The prediction performances are evaluated in terms of MSE, MAE, and R2- score. All the models are trained on an open-source data set associated with HAR. Their performance comparison shows that the LSTM outperforms other neural networks, giving the best prediction accuracy. The authors propose the use of hybrid models and multimodal HAR in future work to improve prediction accuracy. Additionally, to increase the reliability and efficiency of the deep learning models, their time and space complexity in the context of edge devices plays a pivotal role; therefore, it is considered the future scope of this study.

ACKNOWLEDGMENT

The authors express their gratitude to the Director, Centre for Artificial Intelligence & Robotics (CAIR), Defence Research & Development Organization (DRDO), Bangalore for granting permission to publish this research.

REFERENCES

[1] M. P. De Looze, T. Bosch, F. Krause, K. S. Stadler, & L. W. O'sullivan, “Exoskeletons for industrial application and their potential effects on physical work load,” Ergonomics, vol. 59(5), pp. 671-681, 2016.

[2] Olar, M. L., Leba, M., & Risteiu, M. (2021). Exoskeleton-wearable devices. Literature review. In MATEC Web of Conferences (Vol. 342, p. 05005). EDP Sciences.

[3] J. Mai, C. Yi, & Z. Ding, “Human Activity Recognition of Exoskeleton Robot with Supervised Learning Techniques”, 2021.

[4] M. Pesenti, G. Invernizzi, J. Mazzella, M. Bocciolone, A. Pedrocchi, & M. Gandolla, “IMU-based human activity recognition and payload classification for low-back exoskeletons,” Scientific Reports, vol. 13(1), pp. 1184, 2023.

[5] M. Z. Uddin, & A. Soylu, “Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning,” Scientific Reports, vol. 11(1), pp. 16455, 2021.

[6] I. E. Jaramillo, J. G. Jeong, P. R. Lopez, C. H. Lee, D. Y. Kang, T. J. Ha, & T. S Kim, “Real-Time Human Activity Recognition with IMU and Encoder Sensors in Wearable Exoskeleton Robot via Deep Learning Networks,” Sensors, vol. 22(24), pp. 9690, 2022.

[7] F. Attal, S. Mohammed, M. Dedabrishvili, F.,Chamroukhi, L. Oukhellou, & Y. Amirat, “Physical human activity recognition using wearable sensors,” Sensors, vol. 15(12), pp. 31314-31338, 2015.

[8] X. Li, Y. He, & X. Jing, “A survey of deep learning-based human activity recognition in radar,” Remote Sensing, vol. 11(9), pp. 1068 2019.

[9] J. Chen, Y. Sun, & S. Sun, “Improving human activity recognition performance by data fusion and feature engineering,” Sensors, vol. 21(3), pp. 692, 2021.

[10] A. Murad, & J. Y. Pyun, “Deep recurrent neural networks for human activity recognition,” Sensors, vol. 17(11), pp. 2556, 2017.

[11] M. Atikuzzaman, T. R. Rahman, E. Wazed, M. P. Hossain, & M. Z. Islam, “Human activity recognition system from different poses with cnn,” In 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), IEEE, pp. 1-5. December 2020.

[12] R. Bajpai & D. Joshi, “MoveNet: A Deep Neural Network for Joint Profile Prediction Across Variable Walking Speeds and Slopes,” IEEE Transactions On Instrumentation And Measurement, vol. 70, 2021.

[13] K. Taunk, S. De, S. Verma, & A. Swetapadma, “A brief review of nearest neighbor algorithm for learning and classification,” In 2019 international conference on intelligent computing and control systems (ICCS), IEEE, pp. 1255-1260, May 2019.

[14] P. Areekul, T. Senjyu, H. Toyama, & A. Yona, “Notice of violation of ieee publication principles: A hybrid arima and neural network model for short-term price forecasting in deregulated market,” IEEE Transactions on Power Systems, vol. 25(1), pp. 524-530, 2009.

[15] C. J. Li, & T. Y. Huang, “Automatic structure and parameter training methods for modeling of mechanical systems by recurrent neural networks,” Applied Mathematical Modelling, vol. 23(12), pp. 933-944, 1999.

[16] M. Aslam, S. J. Lee, S. H. Khang, & S. Hong, “Two-stage attention over LSTM with Bayesian optimization for day-ahead solar power forecasting,” IEEE Access, vol. 9, pp. 107387-107398, 2021.

[17] H. Hao, F. Yu, & Q. Li, “Soil temperature prediction using convolutional neural network based on ensemble empirical mode decomposition,” IEEE Access, vol. 9, pp. 4084-4096, 2020.