Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 10, No. 2, June 2022, pp. 410~420
ISSN: 2089-3272, DOI: 10.52549/ijeei.v10i2.3475

Single Input Single Head CNN-GRU-LSTM Architecture for Recognition of Human Activities

Updesh Verma[1], Pratibha Tyagi[2], Manpreet Kaur[3]

[1,2,3]Department of Electrical and Instrumentation Engineering, Sant Longowal Institute of Engineering and Technology (sliet), Longowal, Punjab, India

Article Info

Article history:

Received Oct 30, 2021
Revised May 17, 2022
Accepted May 28, 2022

Keyword:

CNN
GRU
LSTM
deep models
neural networks
human activity recognition

ABSTRACT

Due to its applications for the betterment of human life, human activity recognition has attracted more researchers in the recent past. Anticipation of intension behind the motion and behaviour recognition are intensive applications for research inside human activity recognition. Gyroscope, accelerometer, and magnetometer sensors are heavily used to obtain the data in time series for every timestep. The selection of temporal features is required for the successful recognition of human motion primitives. Different data pre-processing and feature extraction techniques were used in most past approaches with the constraint of sufficient domain knowledge. These approaches are heavily dependent on the quality of handcrafted features and are also time-consuming and not generalized. In this paper, a single head deep neural network-based approach with the combination of a convolutional neural network, Gated recurrent unit, and Long Short Term memory is proposed. The raw data from wearable sensors are used with minimum pre-processing steps and without the involvement of any feature extraction method. 93.48 % and 98.51% accuracy are obtained on UCI-HAR and WISDM datasets. This single-head deep neural network-based model shows higher classification performance over other architectures under deep neural networks.

Corresponding Author:
Updesh Verma,
Department of Electrical and Instrumentation Engineering, Sant Longowal Institute of Engineering and Technology (sliet), Longowal, Punjab, India
Email: updesh.verma01@gmail.com

1. INTRODUCTION

Human activity recognition (HAR) has emerged as a very challenging topic nowadays. It has various applications for human beings, such as health assistance, intelligent surveillance, intelligent homes, rehabilitation, abnormal behaviour recognition, personal fitness, gaming, etc. HAR is a process of a group of some steps to recognize human physical motions, posture, and ambulation with the help of various sensors. Different types of sensors have been used previously in HAR, such as video sensors, wearable sensors, ambient sensors, object sensors, smartphone sensors, wireless sensors, etc. In this way, HAR can be broadly classified into two streams one is sensor-based, and the other one is video-based. In the video-based stream of HAR, video sensors like cameras are generally used to take images and videos of human motions and classify them according to the framework[1]. In other streams of sensor-based approaches, wearable sensors like accelerometer, gyroscope, and magnetometer in the form of inertial measurement unit (IMU), ambient sensors like RADAR sensors, wi-fi sensors, and object sensors such as pressure sensors are used for recognition of human activities[2].

Machine learning algorithms have been widely used in previous research in HAR, such as K-Nearest Neighbour (KNN)[3], Support Vector Machine(SVM)[4], and Random Forests[5]. Ensemble empirical mode decompositionbased features and game theory-based feature selection methods were used for activity recognition [6]. The evaluation of these features and the selection method was done by using KNN and SVM

classifiers. Sensor and classifier fusion layer was used in the hierarchical fusion model based on entropy, and for estimation of weights, the weight entropy method was used[7]. Machine learning-based methods require a very time-consuming and qualitative feature extraction technique, feature selection methods, and at the same time, domain knowledge. After consideration of a lot of preparation, pre-processing, and feature extraction operations on data, only decent performance can be achieved. These HAR systems based on machine learning are applicable only for specific applications and cannot classify similar tasks from other sources[8].

Deep learning-based models have gained tremendous momentum nowadays due to their successful performance in the fields of natural language processing, object detection. Image segmentation, classification, etc. Deep learning models gained momentum in HAR over machine learning due to their automatic feature extraction from raw sensor data. These extracted deep features can represent the original data more closely. Deep learning models with minimum pre-processing steps are capable of classifying different labels with minimum human intervention. Different deep learning models such as Deep Belief Network (DBN)[9], Convolutional Neural Network (CNN)[10], Deep feedforward neural networks, and Recurrent Neural Networks (RNN)[11] are heavily used in some of the past research. HAR has been recognized by exploiting some deep learning models such as LSTM, CNN, CNN-LSTM, and more. A smartphone-based activity recognition model based on CNN was proposed in[12]. Time series data of multivariate were utilized and classified with deep CNN framework in[13]. Another deep learning model based on CNN with time-series data of univariate was proposed for an end to end classification in[14]. The accelerometer sensor was used to recognize human activities with CNN-based architecture. CNN was used to extract local features, and the global characteristics of the signal were defined with the help of statistical features[15]. CNN as a multilayer classifier, where CNN and Pooling layers were used alternatively, followed by a fully connected layer for HAR in[16]. This model has experimented on publicly available UCI-HAR datasets [4].

Another neural network that has been widely exploited in the series of deep neural networks is the Recurrent Neural Network (RNN). HAR is a classification problem inside the time series dataset, so that temporal dependency is heavily required for such types of time series datasets for an efficient classification task. RNN provides good temporal dependencies for time series classification problems so that various RNN models were widely used in previous research. For example[17], a framework based on an LSTM feature extractor was proposed to experiment with the WISDM dataset to recognize human activities[18]. In another LSTM model, where the accelerometer and gyroscope sensor's data were first normalized, normalized data was passed through stacked LSTM and then utilized soft-max activation function for classification task[19]. Another bi-directional LSTM for recognizing human activities by accelerometer and gyroscope sensors of mobile phone, mounted on subject's waist was proposed in [20]. Bi-directional LSTM was proposed using the smartphone-based publicly available dataset UCI-HAR to recognize human activities[21]. In recent times various combinations of CNNs layers and RNN layers have been used. For an example, activity recognition system based on CNN followed by RNN dense layer was proposed for HAR in[22]. One model based on the combination of CNN and LSTM was proposed to take advantage of both networks. That model could be able to access data from multimodal sensors without heavy data pr-processing steps. That model utilized gyroscope and accelerometer data in both the cases, individually and combination[23]. Two LSTM layers followed by convolutional layers based model and after the extraction of features that model was succeeded by Global Average Pooling(GAP), Batch normalization and soft-max activation[24].

In recent studies in literature, the multi-head deep neural network with the combination of CNN and RNN has been proposed. For instance, in[25], multi-head CNN-RNN architecture for anomaly detection with multiple sensors was proposed in an industrial environment. In that model, one CNN head was used for one sensor, and then the feature map of each CNN head was concatenated and fed to the next RNN model for finding the temporal information in the feature map. One CNN and one LSTM head were connected in parallel as a multi-head system, and the feature map was then concatenated and passed to the soft-max function for classification in[26]. Multihead LSTM, multi-head CNN-LSTM, and multi-head Conv-LSTM models were designed and ensembled in[27] for recognition of the expenditure of patients on medications.

Some other methods like Extreme Learning Machine(ELM), which is a feedforward network and has no backpropagation ability was proposed for recognition of human motions in[28][29]. Self-adapted architecture with a new sensor location based on ELM was proposed in[30]. The U-net-based model was proposed in[31] for activity recognition by using time series signals of sensors. Most of the approaches defined in past studies for HAR were utilized various feature extraction and selection methods. The accuracy of those approaches depended on the quality of handcrafted features and required expert knowledge of the domain[32]. In this paper, a single-input single head CNN-GRU-LSTM model is designed for HAR using raw wearable sensor data such as gyroscope, accelerometer without the process of handcrafted feature extraction. The feature extraction process inside deep learning models faces significant challenges due to the imbalanced and noisy data obtained from a smartphone or inertial wireless measurement units. In this modelling, the size of the

filter in the convolutional layer is chosen as 7. This combination of three deep neural networks made HAR fruitful in terms of less computational cost, relevant accuracy, and good f-1 score by taking advantage of the trio. CNN extracted the local features, and GRU, LSTM maintains the long-term temporal dependencies of mapped features. The model experimented over two publicly available datasets UCI-HAR[4] and WISDM[33].

1.1. Contribution of this paper

a. Hybrid architecture of deep neural networks by taking the advantages of CNN, GRU, and LSTM is proposed to recognize human activities by taking the raw wearable sensor data with negligible pre-processing steps and without a handcrafted feature extraction process.
b. The local features are extracted by the CNN layer and GRU; LSTM maintains the long-term temporal dependencies of these mapped features so that the model can recognize the diverse data.
c. The model is experimented with over two publicly available datasets UCI-HAR, WISDM and gained accuracy of 93.48 %, 98.51%, respectively.

1.2. Organization of paper

The rest of the paper contains the following sections: Section II describes the methodology behind this proposed work of a single input single head CNN-GRU-LSTM activity recognition model. Section III describes the proposed approach's experiments and results, and the last section IV, consists of the conclusion.

2. METHODOLOGY

Human activity recognition is a time series classification problem. Successfully detecting activities from raw sensor data obtained from smartphone and wearable units requires an efficient feature extraction process. This proposed single input single head CNN-GRU-LSTM model extracts temporal feature map from the CNN layer then long-term temporal dependencies are maintained by GRU and LSTM layers. LSTM layer is also used to eradicate the gradient vanishing and exploding problem, which is common in CNN. This deep neural network approach finds the way for end-to-end classification from raw sensor data to feature extraction and then classification—the architecture experiments on two publicly available datasets UCI-HAR and WISDM.

2.1. Data Segmentation

The first step towards activity recognition is a segmentation of acquired raw data from wearable sensors. This process of data segmentation is carried out with the help sliding window. WISDM data set is segmented in the timesteps of 128 and 3 features for every timestep. UCI -HAR data was already available in a segmented form where data was segmented into 128 timesteps, and every timestep consisted of 9 features. So that the input vector size for the WISDM dataset is 128x3 and the input vector size for the UCI-HAR dataset is 128x9. This 128-size input vector is considered as one sample for one activity. This vector size of length 128 is calculated over n channels where the value of n is equal to the value of features of a particular dataset. In this case, the value of n for the WISDM dataset is three and for the UCI-HAR dataset is 9.

2.2. Feature Extraction

The advantages of both CNN and RNN are utilized by combining CNN and RNN. These neural networks have the capability of automatic feature extraction. CNNs[34] are generally used for processing the data of multiple arrays.

The architecture of CNN generally consists of CNN layer, pooling, and at last, fully connected layer. Operation of convolution on time series data of length K and width M is depicted in fig1., where M is nothing but the features available in the dataset. Feature map is generated after convolution between filters of length n and depth h, and time-series data of length K and features M. Each convolution unit generates its feature map. A set of local input time series data and kernel of size n x h are multiplied by exact overlapping with each other so that the size of the regional input time series, which is also called the receptive field, must be equal to the size of the filter for the generation of the feature map. Each value in the receptive field is multiplied with the weights of the filter bank, and then all the obtained values are summed up then obtained one number is

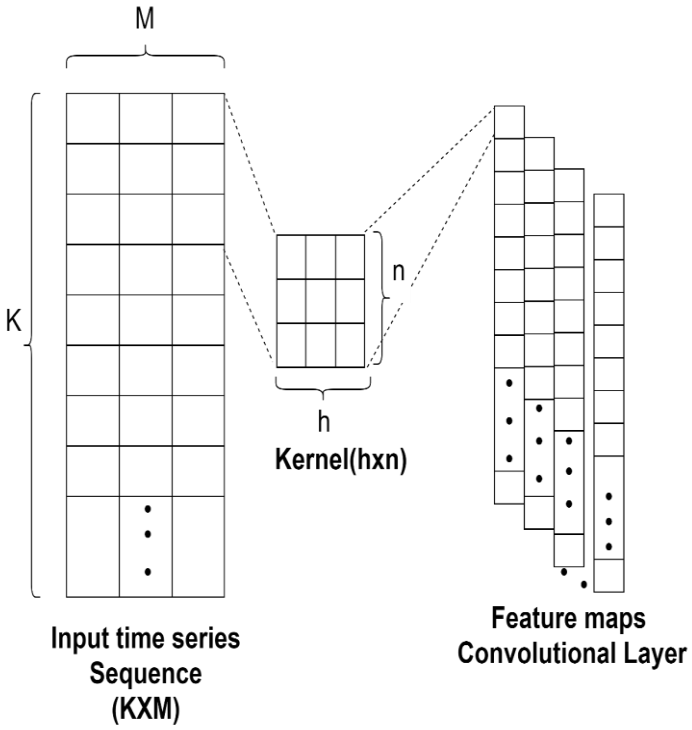


Figure 1. Operation of Convolutional layer on input data

The single value of the feature map. This value obtained by multiplication and addition is then passed through the non-linear function called activation function Rectified Linear Unit. The number of feature maps depends on the number of kernels used. These feature maps then passed through the pooling layer, which helped reduce the feature map's dimension by taking the maximum value inside the local patches. A regularly used Dropout layer reduces the chances of overfitting.

CNN performance is recognized as the extraction of local features by taking time-series data in frames. Local values inside the data frame are highly correlated with concerned activities. CNN takes each frame of the dataset independently and takes this information in terms of a feature map, so it can be said that CNN follows the shortterm dependencies. So, it is crucial to extract the features locally due to the high correlation between the values of features for activity recognition; hence CNN performs the crucial task. But long-term dependencies are also required for precise recognition of activities.\

RNN is introduced to gain the advantages of long-term dependencies, but due to the large size of the activity dataset, the gradient vanishing problem[35] gets associated with traditional RNN. So traditional RNN is not useful for activity recognition. GRU is introduced[36] to overcome this gradient vanishing and gradient exploding problem in traditional RNN. GRU, an extension of traditional RNN, is used to gain the advantages of long-term dependencies of time series sequences for the activity dataset[37].

GRU and LSTM layers are added after the CNN layer to connect the past information with the present scenario in the proposed model. The reset gate and update gate is part of the GRU unit, and LSTM consists of the input gate, output gate, and forget gate. LSTM can be able to capture more long-term dependencies than the GRU unit. The human activities dataset is generally considered as a very long dataset in dimensions so that intense temporal longterm dependencies are required. Both unit LSTM and GRU in combination can provide very long-term dependencies, and the model with these combinations can handle the data of large variations and diversity.

2.3. Proposed Architecture

In the following fig 2, the different parts of the proposed model are shown in the form of part A, part B, and Part C. Part A describes the CNN unit this unit followed by the GRU unit and LSTM unit, which are described as part B and part C respectively. In part, three convolutional 1D layers are used with the activation function of Relu and each CNN layer, followed by max-pooling and dropout layer. In part B, three GRU units are used, and a dropout layer follows each GRU unit. In part C first LSTM layer and second LSTM layer are followed by the dropout layer and flatten layer, respectively.

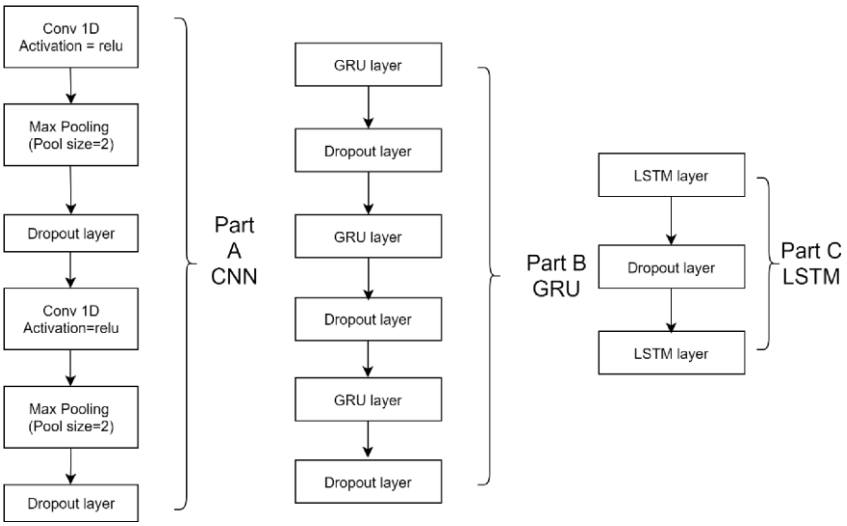


Figure 2. Different parts of CNN-GRU-LSTM

The complete architecture of the proposed model with the combination of part A, part B, and Part C is shown in fig 3. 64 filters in the first layer of CNN and 32 filters in the remaining two layers are used for extraction of local features. A 10% dropout follows each layer of CNN with a filter size of 7. 10% dropout is selected by the number of experiments in the range of 10% to 50% dropout and the observation of impact on accuracy. Three GRU layers with 64 units in the first layer and 32 units in the remaining two layers are used in part B to capture the long-term dependencies and return sequences passes to the next layers. A 10% dropout follows each GRU unit. Two LSTM layers with 32 units are used in part C, where the first layer is followed by 10% dropout, and a flattening layer follows the second. LSTM layers are used to understand better more temporal long-term dependencies of time series sequences of data.

In the last section flatten layer is followed by a dense layer with 128 units, and this, in turn, is followed by a 10% dropout. The Batch-normalization layer is used for the normalization of the feature map for better classification efficiency. At last, a fully connected layer with a soft-max activation function is used.

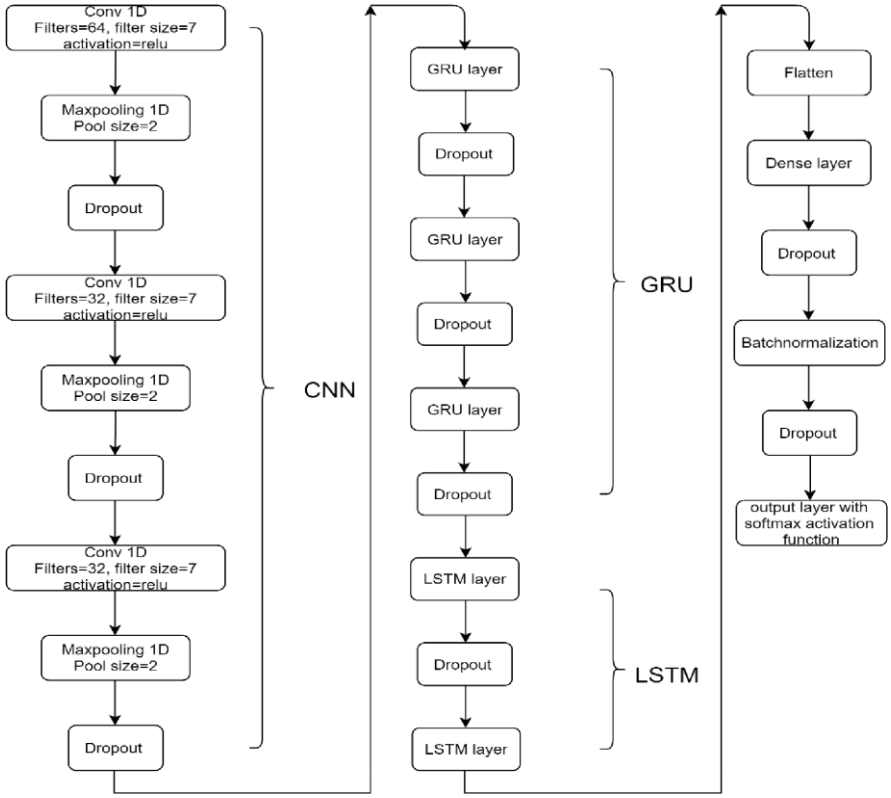


Figure 3 proposed Architecture CNN-GRU-LSTM for human activity recognition

3. EXPERIMENTS AND RESULTS

The proposed model is trained and tested on UCI-HAR and WISDM datasets. Table 1 describes the hyperparameters for the proposed model. Other hyperparameters are taken as default. The model is trained to reduce the loss function. Keras with TensorFlow backend is used for implementation of the proposed model in google colab. The values of parameters for performance evaluation, datasets used, and results are described in this section.

Table 1. description of hyperparameters

Hyperparameter	Used values
Loss function	Sparse categorical cross-entropy
Optimizer	Adam
Batch size	400
Learning rate	0.001
Dropout	10%
Filter size	7
Pool size	2
Length of input vector	128
Epochs	500
Number of channels	9 for UCI-HAR dataset 3 for the WISDM dataset

3.1. Datasets

UCI-HAR[4]: The University of California Irvine (UCI) released the dataset and publicly available it on the UCI repository. Thirty volunteers participated, and waist-mounted smartphone's inbuilt sensors like accelerometer and gyroscope were used. Total six human activities (sitting, standing, walking, walking downstairs, and walking upstairs) were recorded with the sampling frequency of 50Hz. Three axial body acceleration, three axial angular velocities, and a total of three axial acceleration; hence a total of 9 features were measured and stored. Butterworth low pass filter with a cut-off frequency of 0.3 Hz was used to segregate body acceleration and gravity. 2.56 s window was used for data segmentation, and 70% of the total subject's data were recorded as training data and the rest as testing data. A total of 10299 samples were recorded, where 7352 were taken as training samples and 2947 as testing samples.

WISDM[33]: Wireless sensor data mining lab of Fordham University was used to acquire the WISDM dataset. A total of 36 volunteers participated in that project, and they performed six activities (jogging, descending stairs, ascending stairs, walking, sitting, and standing) with placing a smartphone in their pocket. An inbuilt accelerometer sensor of the smartphone with a sampling frequency of 20Hz was used for acquiring the data. Twenty-nine subjects are selected for this paper to train the proposed model and the rest for validation. Data normalization is performed with the normalization of all values in the range of 0-1.

3.2. Performance Representation

The performance of the proposed model is represented by accuracy score, f1-score, precision, recall, and confusion matrix. Accuracy is defined as the ratio of correctly classified labels or targets to a total number of samples. Correctly classified labels are generally known as adding a total number of true positives (TP) and true negatives (TP). Wrongly classified terms are generally considered as false positive (FP) and false-negative (FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

The ratio of samples that are correctly predicted positive to the all-positive predicted samples is known as precision.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

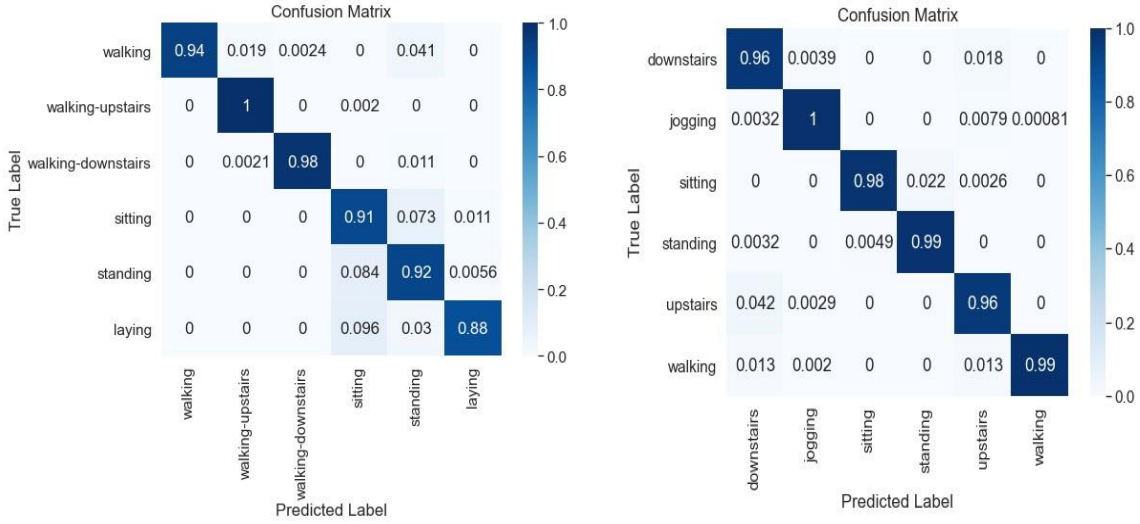
The ratio of correctly predicted positives to the samples which actually exist as positive, known as recall.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

For the performance evaluation of the model that is going to test on an unbalanced dataset, the f1-score is generally calculated. The harmonic mean of precision and recall is known as the f1 score.

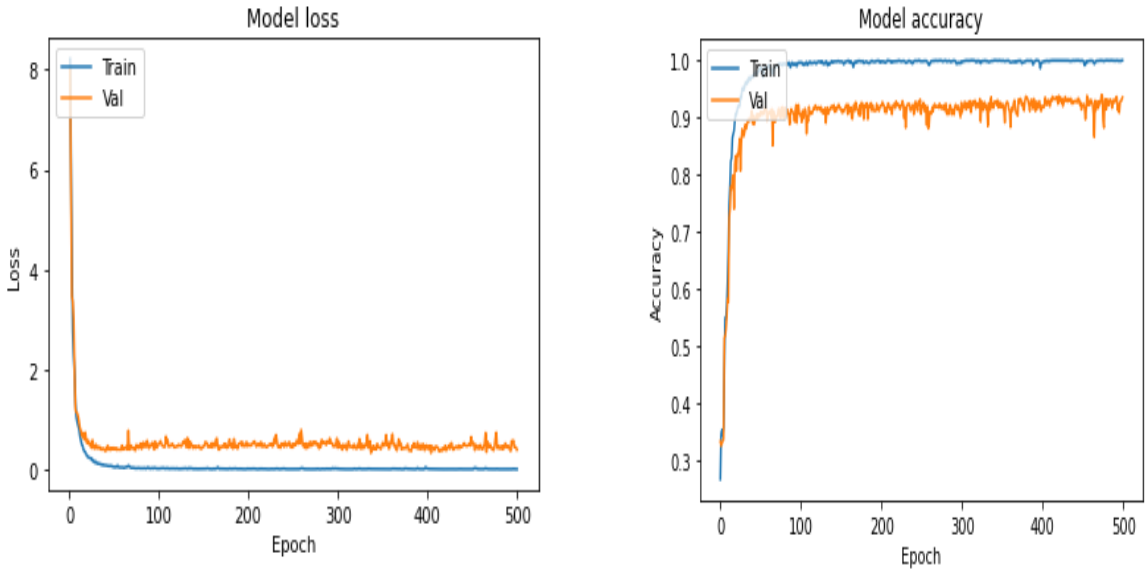
$$f1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (4)$$

The confusion matrix describes all the predicted classes and actual classes with their interrelationship positively or negatively. Rows of this matrix are the actual classes, and columns are considered as predicted classes.



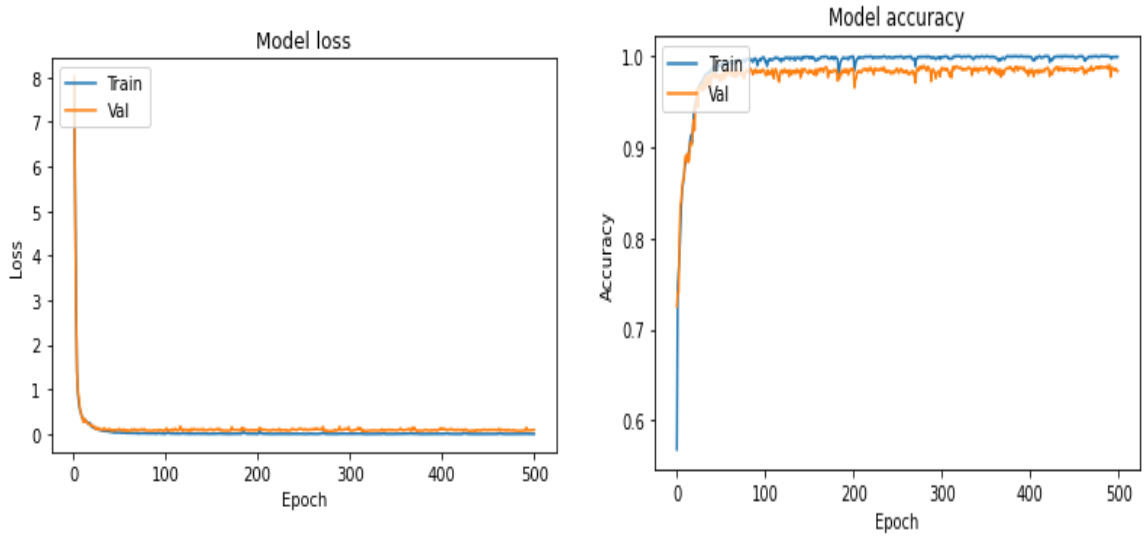
(a) (b)

Figure 4. (a) Confusion matrix of the proposed model on UCI-HAR dataset (b) confusion matrix of the proposed model on WISDM dataset



(a) (b)

Figure 5. (a) Plot of training and testing loss of proposed model on UCI-HAR dataset (b) Plot of training and testing accuracy of proposed model on UCI-HAR dataset.



(a) (b)

Figure 6. (a) Plot of training and testing loss of proposed model on WISDM dataset (b) plot of training and testing accuracy of the proposed model on WISDM dataset.

3.3. Results and Discussion

Fig 4(a) shows the confusion matrix of the proposed CNN-GRU-LSTM architecture on the test dataset of UCI HAR and fig 4(b) shows the confusion matrix on the test dataset WISDM of the proposed model. Both the figures under fig. 4 describe that the five activities of the test dataset of the UCI-HAR dataset obtained accuracy and f1-score greater than 90% and all the activities of the test dataset of the WISDM dataset obtained accuracy and f1-score greater than 95%. Fig 5(a) shows the training and testing loss of the proposed model on the UCI-HAR dataset and fig 5(b) shows the training and testing accuracy of the proposed model on the UCI-HAR dataset. In the same way, fig 6(a) and fig 6(b) show the training and testing loss of the proposed model on the WISDM dataset and the training and testing accuracy of the proposed model on the WISDM dataset, respectively. Table 2 and Table 3 describes the effectiveness of proposed model on both the dataset. The effectiveness of the proposed model on the UCI-HAR dataset is described in table 2, where some previous models are taken for comparison with the proposed model. It is found that the proposed model outperformed six models in terms of accuracy and f1-score on the UCI-HAR dataset.

The effectiveness of the proposed model on the WISDM dataset is described in table 3, where some previous models are taken for comparison with the proposed model. It is found that the proposed model outperformed nine models in terms of accuracy and f1-score.

Table 2. Comparative performance of proposed model on UCI-HAR dataset

Frameworks	Accuracy (%)	F-1 score (%)
Res-LSTM [38]	91.6	91.5
CNN-LSTM [39]	92.13	-
Stacked LSTM [40]	93.13	-
CNN [41]	92.71	92.93
Single input CNN-GRU model A[42]	93.03	93.01
Single input CNN-GRU model B[42]	92.43	92.42
Proposed	93.48	93.5

Table 3. Comparative performance of frameworks on WISDM dataset

Frameworks	Accuracy (%)	F-1 score
Statistical features and reweighted genetic algorithms [43]	94.02	-
CNN [44]	93.32	-
LSTM-CNN [45]	95.85	-
U-Net [46]	96.4	96.5
Single input CNN-GRU model A [42]	92.03	92.42
Single input CNN-GRU model B [42]	94.71	94.50
Single input CNN-GRU model C [42]	92.37	92.55
Multi input CNN-LSTM [42]	95.54	95.55
Multi input CNN-GRU [42]	97.21	97.22
Proposed	98.51	98.52

The novelty of the proposed model is to introduce the in-depth knowledge of long-term dependencies to the model by using both the units GRU and LSTM. Comparative analysis with similar research is included in this paper and found that this less complex model performed better than some existing research. Outstanding performance is observed in the case of the WISDM dataset with an accuracy of 98.51% and an f1-score of 98.52%. Due to the well understanding of short-term and long-term dependencies on temporal sequences of time series datasets, architecture could handle the diverse nature of data.

Table 4. Computational Performance analysis on the basis of trainable parameters

Model	UCI-HAR dataset			WISDM dataset		
	Parameters	Accuracy (%)	F-1 score (%)	Parameters	Accuracy (%)	F-1 score (%)
Standard CNN [47]	-	-	-	1.55M	96.83%	-
Residual Network [47]	-	-	-	2.30M	98.32%	-
Zhang et al [48]	-	-	-	2.77M	96.4	95.4
2D CNN (without total acceleration) [49]	7.314M	85	-	-	-	-
2D CNN (with total acceleration) [49]	7.314M	85.2	-	-	--	-
Proposed	1.13M	93.48	93.5	1.13M	98.51	98.52

Table 4 represents the computational efficiency of proposed architecture over some previous researches on the basis of used number of trainable parameters. Standard CNN was implemented in [47] by Gao et al., with 1.55 million parameters and obtained lesser accuracy by 1.68% than our proposed work with 1.13million parameters on WISDM dataset. The residual network was implemented in [47] with 2.30 million trainable parameters and obtained a lesser accuracy by 0.19% than our model with 1.13 million trainable parameters on same dataset. The Multi-head attention-based CNN model were used in [48] for HAR and achieved a lesser accuracy and F-1 score by 2.11% and 3.12% respectively with 2.77 million trainable parameters on WISDM dataset than our proposed architecture. The various deep learning models were designed by Tufek et al., in [49] on UCI-HAR dataset and two conditions were applied on dataset. In the first condition only accelerometer and gyroscope were considered and in second condition, the whole dataset was used including total acceleration and it is observed that our model gets much higher accuracy with fewer number of parameters. Hence, we can say that as far as computational efficiency is concerned then this research, proposed in this paper performed well than the other similar previous researches.

4. CONCLUSION

Convolutional Neural Networks and Recurrent Neural Networks for human activity recognition are implemented and tested on publicly available datasets. The framework CNN-GRU-LSTM outperformed some similar research in this field. These architectures gained advantages of three neural networks CNN as well as GRU and LSTM. CNN generates local features of the input sequence with efficient local

dependencies, and GRU utilizes its efficiency of capturing the long-term dependencies. Long-term dependencies in depth are provided to this framework by the LSTM unit, and hence this proposed model can handle more diverse data. The performance of the proposed model is tested on publicly available datasets such as WISDM and UCIHAR, and it is found that the architecture outperformed some multi-head architectures. This single head and input model with diverse data handling capacity is computationally efficient and more accurate than other similar architectures.

REFERENCES

[1] Z. A. Khan and W. Sohn, "A hierarchical abnormal human activity recognition system based on R-transform and kernel discriminant analysis for elderly health care," *Computing*, vol. 95, no. 2, pp. 109–127, 2013, doi: 10.1007/s00607-012-0216-x.

[2] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "Using Wearable Sensors," vol. 17, no. 2, pp. 386–403, 2017.

[3] A. Jain and V. Kanhangad, "Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors," *IEEE Sens. J.*, vol. 18, no. 3, pp. 1169–1177, 2018, doi: 10.1109/JSEN.2017.2782492.

[4] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones.," in *Esann*, 2013, vol. 3, p. 3.

[5] Z. Feng, L. Mo, and M. Li, "A Random Forest-based ensemble method for activity recognition," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-Novem, pp. 5074–5077, 2015, doi: 10.1109/EMBC.2015.7319532.

[6] Z. Wang, D. Wu, J. Chen, A. Ghoneim, and M. A. Hossain, "A Triaxial Accelerometer-Based Human Activity Recognition via EEMD-Based Features and Game-Theory-Based Feature Selection," *IEEE Sens. J.*, vol. 16, no. 9, pp. 3198–3207, 2016, doi: 10.1109/JSEN.2016.2519679.

[7] M. Guo, Z. Wang, N. Yang, Z. Li, and T. An, "A multisensor multiclassifier hierarchical fusion model based on entropy weight for human activity recognition using wearable inertial sensors," *IEEE Trans. Human-Machine Syst.*, vol. 49, no. 1, pp. 105–111, 2019, doi: 10.1109/THMS.2018.2884717.

[8] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, 2018, doi: 10.1016/j.eswa.2018.03.056.

[9] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Commun. ACM*, vol. 54, no. 10, pp. 95–103, 2011, doi: 10.1145/2001269.2001295.

[10] A. Carruthers and J. Carruthers, "Introduction.," *Dermatol. Surg.*, vol. 39, no. 1 Pt 2, p. 149, 2013, doi: 10.1111/dsu.12130.

[11] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," pp. 1–9, 2014.

[12] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep Learning Models for Real-time Human Activity Recognition with Smartphones," *Mob. Networks Appl.*, vol. 25, no. 2, pp. 743–755, 2020, doi: 10.1007/s11036-019-01445-x.

[13] C. L. Liu, W. H. Hsaio, and Y. C. Tu, "Time Series Classification with Multivariate Convolutional Neural Network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4788–4797, 2019, doi: 10.1109/TIE.2018.2864702.

[14] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *J. Syst. Eng. Electron.*, vol. 28, no. 1, pp. 162–169, 2017, doi: 10.21629/JSEE.2017.01.18.

[15] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," *Appl. Soft Comput. J.*, vol. 62, pp. 915–922, 2018, doi: 10.1016/j.asoc.2017.09.027.

[16] C. A. Ronao and S. B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, 2016, doi: 10.1016/j.eswa.2016.04.032.

[17] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM Networks for Mobile Human Activity Recognition," no. Icaita, pp. 50– 53, 2016, doi: 10.2991/icaita-16.2016.13.

[18] L. Yu, J. Shao, X. S. Xu, and H. T. Shen, "Max-margin adaptive model for complex video pattern recognition," *Multimed. Tools Appl.*, vol. 74, no. 2, pp. 505–521, 2015, doi: 10.1007/s11042-014-2010-6.

[19] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, "Stacked Lstm Network for Human Activity Recognition Using Smartphone Data," *Proc. - Eur. Work. Vis. Inf. Process. EUVIP*, vol. 2019-October, pp. 175–180, 2019, doi: 10.1109/EUVIP47703.2019.8946180.

[20] S. Yu and L. Qin, "Human activity recognition with smartphone inertial sensors using bidir-LSTM networks," *Proc. - 2018 3rd Int. Conf. Mech. Control Comput. Eng. ICMCCE 2018*, pp. 219–224, 2018, doi: 10.1109/ICMCCE.2018.00052.

[21] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors," *Math. Probl. Eng.*, vol. 2018, 2018, doi: 10.1155/2018/7316954.

[22] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan, "InnoHAR: A deep neural network for complex human activity recognition," *IEEE Access*, vol. 7, pp. 9893–9902, 2019, doi: 10.1109/ACCESS.2018.2890675.

[23] R. Mutegeki and D. S. Han, "A CNN-LSTM Approach to Human Activity Recognition," *2020 Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2020*, pp. 362–366, 2020, doi: 10.1109/ICAIIC48513.2020.9065078.
[24] K. Xia, J. Huang, and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," *IEEE Access*, vol. 8, pp. 56855– 56866, 2020, doi: 10.1109/ACCESS.2020.2982225.
[25] M. Canizo, I. Triguero, A. Conde, and E. Onieva, "Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study," *Neurocomputing*, vol. 363, pp. 246–260, 2019, doi: 10.1016/j.neucom.2019.07.034.
[26] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM Fully Convolutional Networks for Time Series Classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017, doi: 10.1109/ACCESS.2017.2779939.
[27] S. Kaushik, A. Choudhury, N. Dasgupta, S. Natarajan, L. A. Pickett, and V. Dutt, "Ensemble of Multi-headed Machine Learning Architectures for Time-Series," *Appl. Mach. Learn.*, p. 199, 2020.
[28] V. B. Semwal, N. Gaud, and G. C. Nandi, "Human gait state prediction using cellular automata and classification using ELM," in *Machine intelligence and signal analysis*, Springer, 2019, pp. 135–145.
[29] P. Patil, K. S. Kumar, N. Gaud, and V. B. Semwal, "Clinical Human Gait Classification: Extreme Learning Machine Approach," *1st Int. Conf. Adv. Sci. Eng. Robot. Technol. 2019, ICASERT 2019*, vol. 2019, no. Icasert, 2019, doi: 10.1109/ICASERT.2019.8934463.
[30] Z. Wang, D. Wu, R. Gravina, G. Fortino, Y. Jiang, and K. Tang, "Kernel fusion based extreme learning machine for cross-location activity recognition," *Inf. Fusion*, vol. 37, pp. 1–9, 2017, doi: 10.1016/j.inffus.2017.01.004.
[31] M. A. K. Quaid and A. Jalal, "Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm," *Multimed. Tools Appl.*, vol. 79, no. 9–10, pp. 6061–6083, 2020, doi: 10.1007/s11042-019-08463-7.
[32] J. Lu, X. Zheng, M. Sheng, J. Jin, and S. Yu, "Efficient Human Activity Recognition Using a Single Wearable Sensor," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11137–11146, 2020, doi: 10.1109/JIOT.2020.2995940.
[33] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, 2011.
[34] Y. LeCun, Y. Bengio, and others, "Convolutional networks for images, speech, and time series," *Handb. brain theory neural networks*, vol. 3361, no. 10, p. 1995, 1995.
[35] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. neural networks*, vol. 5, no. 2, pp. 157–166, 1994.
[36] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv Prepr. arXiv1409.1259*, 2014.
[37] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv Prepr. arXiv1412.3555*, 2014.
[38] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep residual bidir-LSTM for human activity recognition using wearable sensors," *Math. Probl. Eng.*, vol. 2018, 2018.
[39] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 362–366.
[40] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, "Stacked lstm network for human activity recognition using smartphone data," in *2019 8th European workshop on visual information processing (EUVIP)*, 2019, pp. 175–180.
[41] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mob. Networks Appl.*, vol. 25, no. 2, pp. 743–755, 2020.
[42] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors," *Computing*, pp. 1–18, 2021.
[43] M. A. K. Quaid and A. Jalal, "Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm," *Multimed. Tools Appl.*, vol. 79, no. 9, pp. 6061–6083, 2020.
[44] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," *Appl. Soft Comput.*, vol. 62, pp. 915–922, 2018.
[45] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020.
[46] Y. Zhang, Z. Zhang, Y. Zhang, J. Bao, Y. Zhang, and H. Deng, "Human activity recognition based on motion sensor using u-net," *IEEE Access*, vol. 7, pp. 75213–75226, 2019.
[47] Gao, W., Zhang, L., Teng, Q., He, J., & Wu, H. (2021). DanHAR: Dual attention network for multimodal human activity recognition using wearable sensors. *Applied Soft Computing*, *111*, 107728.
[48] H. Zhang, Z. Xiao, J. Wang, F. Li and E. Szczerbicki, "A Novel IoT-Perceptive Human Activity Recognition (HAR) Approach Using Multihead Convolutional Attention," in IEEE Internet of Things Journal, vol. 7, no. 2, pp. 1072-1080, Feb. 2020, doi: 10.1109/JIOT.2019.2949715.
[49] Tufek, N., Yalcin, M., Altintas, M., Kalaoglu, F., Li, Y., & Bahadir, S. K. (2019). Human action recognition using deep learning methods on limited sensory data. *IEEE Sensors Journal*, *20*(6), 3101-3112.