Identifying Activity of Humans using CNN with Auto-Encoders and LSTM

Dr. Gayathri Tata
Assistant Professor, Computer Science and Engineering
Shri Vishnu Engineering College for Women(A)
Bhimavaram, Andhra Pradesh
gayathritcse@svecw.edu.in

Kuriseti Lovely Srenika
Computer Science and Engineering
Shri Vishnu Engineering College for Women(A)
Bhimavaram, Andhra Pradesh
srenu1680@gmail.com

Mogathadakala Alekhya
Computer Science and Engineering
Shri Vishnu Engineering College for Women(A)
Bhimavaram, Andhra Pradesh
alekhya.dtl@gmail.com

Viswanadham Harshita Sai
Computer Science and Engineering
Shri Vishnu Engineering College for Women(A)
Bhimavaram, Andhra Pradesh
Viswanadham.harshitasai@gmail.com

Nallamalli Lasya
Computer Science and Engineering
Shri Vishnu Engineering College for Women(A)
Bhimavaram, Andhra Pradesh
lasyanallamalli27@gmail.com

Nallamolu Tulasi
Computer Science and Engineering
Shri Vishnu Engineering College for Women(A)
Bhimavaram, Andhra Pradesh
tulsinallamolu@gmail.com

Abstract- Recognizing a person's activities has been a critical time series classification problem that includes predicting Human movements based on sensor data. The conventional approach using machine learning algorithms requires in-depth domain knowledge for signal processing to the correct construction of features from raw data. This paper proposes deep learning approaches for human activity recognition with data from sensors such as accelerometers and gyroscopes. The convolution neural networks(CNNs) with ability of automatic feature extraction, autoencoders(AEs) used for dimensionality reduction, and long short-term memory(LSTM) which are good at temporal modeling, complement each other. In this work, we would like to leverage the complementarity of CNNs, AEs and LSTMs by integrating them into a unified architecture.

Keywords— Deep Learning, Convolution Neural Networks, Long short term memory (LSTM), Human Activity Recognition (HAR), Convolutional Layer, Maxpool Layer, Deconvolutional Layer .

I. Introduction

Efficient, Accurate, and Faster Human activity recognition (HAR) can have a wide range of applications. For Instance, in healthcare, particularly in the case of elderly people, HAR, when used with other technologies such as the Internet of Things (IoT), can be leveraged to initiate a quick response to any contingency as the Human activity is being observed, thereby reducing loss of lives. HAR involves recognizing a variety of human activities, including walking, running, sitting, sleeping, standing, showering, cooking, driving, and unusual activities. We can use wearable sensors or accelerometers and video frames to collect data. The crime rate can also be monitored by HAR, and recognizing human activity during vehicle driving can also lead to safe travel.

In this project, we would like to propose and build advanced methodologies and compare them with the previous methods. To build a robust system, we would like to consider datasets where data is collected from various sensors, images, accelerometers, and gyroscopes. We will study various deep neural network architectures based on convolutional neural networks, recurrent neural networks. to solve this problem and propose a novel architecture that can better perform the task.

The primary public dataset that will be used in this project is the "Human Activity Recognition Using Smartphones" dataset introduced by Davide Anguita et al. To collect the data; The experiment was conducted in a group of 30 volunteers within the 19 - 48 years age group. Each person wore a smartphone on their waist and performed six different activities (walking, walking_downstairs, walking_upstairs, standing, sitting, laying). With a built-in accelerometer and gyroscope in the smartphone, recorded 3-axis linear acceleration and 3-axis angular velocity at a constant frequency of 50Hz. The experiment was recorded on video for manual labeling of the data.

II. Literature survey

Many research works have been carried out for recognizing human activities[[1],[2],[3]] using various Machine Learning and Deep Learning approaches. Numerous shallow ML methods have been proposed emphasizing only the accuracy of various ML algorithms[[4],[5]]

The authors in the paper[14] used three publicly available datasets: Skoda[6], OPPORTUNITY[7], and Actitracker[8] to prove the efficiency of their proposed work.

Another CNN model for HAR is proposed by Ronao and Cho[9] using a publicly available UCI dataset[10] and handcrafting the features.

A HAR framework proposed by Ravi et al.[11] used Convolutional layers and shallow features gathered from smartphone and body-worn sensors; WISDM[12] and DAPHNet-FoG[13] datasets are used in this paper.

Jiang et al.[14] employed Adam hyper-parameter optimizer in the CNN model built for HAR; UCI dataset[10] was used.

Another methodology presented by I Andrey[15] used a combination of handcrafted and automatically extracted features from a CNN; the datasets used are WISDM[12] and UCI[10].

Wan et al. [16] presented a methodology for HAR leveraging three different Deep Learning architectures, which are CNN, LSTM, and bidirectional LSTM.

The data is gathered from UCI[10] and PAMAP2[17] datasets, where the latter dataset is gathered from wearable sensor data.

Deep learning Methods such as CNN, LSTM and other combinations resulted in the test accuracy of around 90% with a validation accuracy of around 95% in the previous research works.

ML Algorithms used	Accuracy
Logistic regression	85.83%
Linear SVC	86.74%
Decision Tree	87.78%
Random Forest	90.3%
rbf SVM Classifier	86.27%

Fig 2.1 Performance of some deep learning algorithms

III. SYSTEM DESIGN

Introduction:

System Design can be best understood from design goals and system architecture. Each and every system has its own system goals and significance.

Design goals:

Few system design goals are as follows:

1.Performance requirement: All data entered shall be up to mark and no flaws shall be there for good performance.

2.Platform constraints: The main target is to generate an intelligent system to recognize human activities.

3.Accuracy and Precision: Requirements are accuracy and precision for the data given as input, as well as produced as output.

4.Modifiability: Requirements about the effort needed to make changes in the software. (Person- months).

5.Portability: It is the ability to carry the system whenever needed.

6.Reliability: Requirements about how often the software fails.

7.Security: One or more requirements about protection of your system and its data.

8.Usability: Requirements about how difficult it will be to learn and operate the system.

System Architecture:

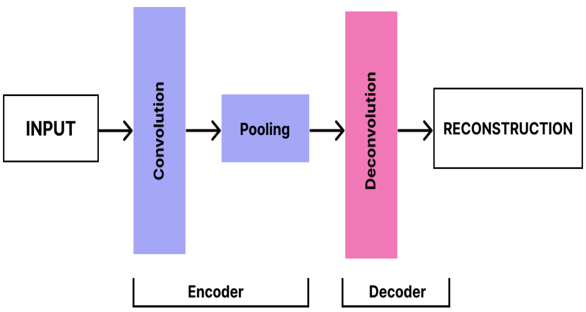

Fig 3.1 System Architecture

The model[Fig 3.1] proposed consists of three modules, the first being the convolutional AE consisting of a convolutional layer, a pooling layer and a deconvolutional layer. The output of this layer is passed through a flattened layer which serves as a desired input for the LSTM layer. The LSTM output goes through a fully connected layer to get a high level representation.

The proposed Convolutional AE comprises convolution, pooling and deconvolution layers as shown in Fig 3.1. Encoder includes a convolutional layer and a Maxpool layer whereas Decoder comprises a deconvolutional layer. The results from the convolutional layer are encoded with a Maxpool layer that allows high-layer representations that doesn't alter with small changes in the inputs which reduces the overall computational cost.

Temporal features play an important role in modeling human movements. In recent years, LSTMs have performed impressively in HAR and various other domains. The temporal features are extracted from time sensory signals by the LSTM architecture because of its long-term dependencies and temporal characteristics. In our proposed architecture as explained earlier, convolutional AE is followed by a LSTM model. The results of the convolutional AE are passed as inputs to the LSTM to deduce the latent temporal interactions across the timeframes.

IV. Implementation

Implementation:

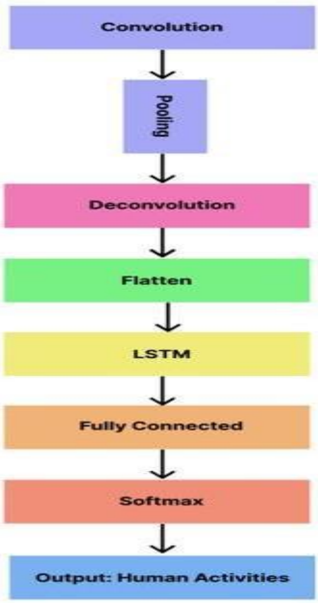

Fig 4.1 Implementation

The network[Fig 4.1] is trained using the training set where the error is being calculated by taking the actual output and the predicted output. The errors are then back propagated using the Adam optimizer in sequence of the layers to update the hyperparameters of the network.

To perform the experiment, the first two datasets are divided into two different groups. 70% of volunteers are selected for

training and 30% are used to test the proposed HAR solution. Therefore, data from the same subject is not included in both the training and test datasets.

In our experiment, we use a simple 5-way cross-validation test to generate multiple training and validation splits from a training set[Fig 4.2]. This is because cross-validation is less computationally complex than other methods such as leave-one-out cross validation.

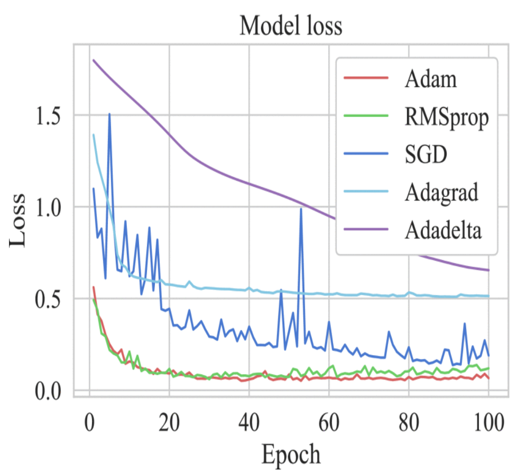

Fig 4.2 Model Loss

We have used the data from one subject for the test and the data from the other subject for training. This cross-subject test is more rigorous because the test data is hidden from the model, making it a more realistic framework for verifying the generalization capabilities of the model. All datasets are used to perform 1D convolution on the input layer of the CNN. In our experiments, ReLU is used as the activation function for a convolutional layer with a kernel size of 3, a stride of 2, and a filter size of 64. The learning rate (alpha) is set to 0.001. The optimizer reduces the loss function by updating and calculating network parameters that affect the training and output process of the model to approach or reach optimal values.

With CNN, the training and test calculation times are 3.4374 seconds and 371.6 ms, respectively. Using the UCI dataset, the test accuracy of Conv-AE with LSTM is 93.42%, which is much higher than other common DL approaches.

[Fig [4.3],[5.1]] shows the detailed classification results of the proposed model. This proposed model uses both convolutional AEs and LSTMs in combination, so the F1 values for activities such as walking, stairs down, and stairs up are 99%, 100%, and 94%, respectively. Therefore, it can be concluded that the proposed method can identify similar activity patterns very efficiently.

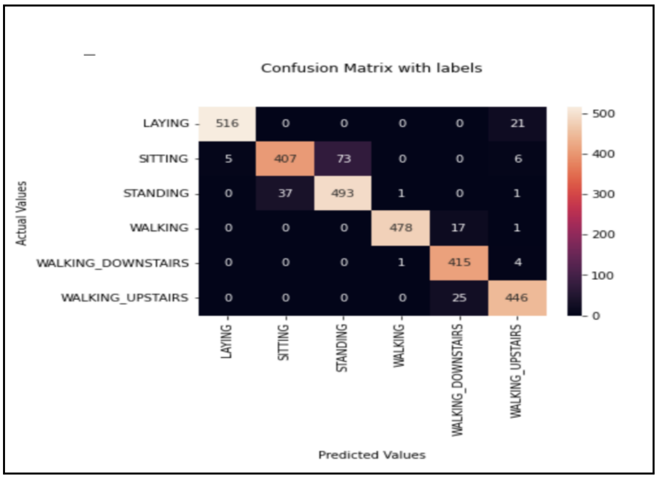

Fig 4.3 Confusion Matrix

V. RESULTS

Model predicts the activity of a person by 95-99%[Fig [5.2],[5.3]] depending on the position of the person.

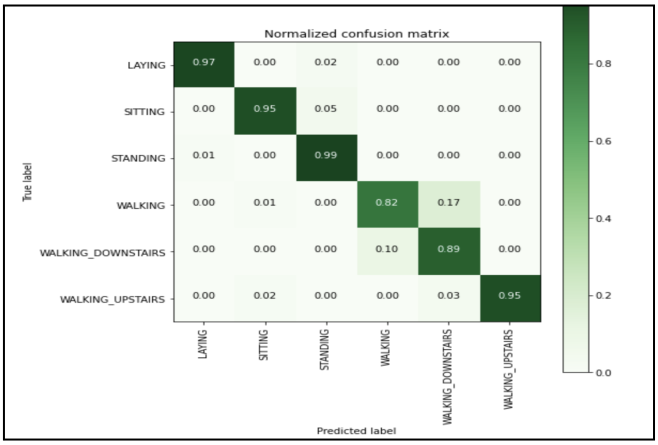

Fig 5.1 Normalized Confusion Matrix

	precision	recall	f1-score	support
0	0.99	0.97	0.98	496
1	0.97	0.95	0.96	471
2	0.92	0.99	0.95	420
3	0.88	0.82	0.85	491
4	0.82	0.89	0.86	532
5	1.00	0.95	0.97	537
accuracy			0.93	2947
macro avg	0.93	0.93	0.93	2947
weighted avg	0.93	0.93	0.93	2947

Fig 5.2 Performance

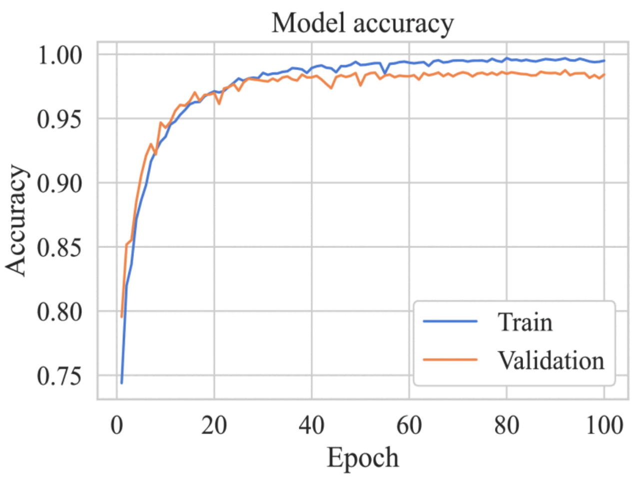

Fig 5.3 Model accuracy

VI. CONCLUSION

In conclusion, we proposed a unified architecture. This architecture combines the abilities of CNN, AE and LSTM. CNN with its automatic feature extraction and AE with its ability of efficient dimensionality reduction and LSTM with the ability to preserve sequence information.

Currently we have primarily got the performance results of the proposed architecture on UCI-HAR dataset.
Further, we want to study the performance of the model in other datasets for this research work to be presented.

VI. References

[1] D.Anguita,A.Ghio,L.Oneto,X.Parra,andJ.L.Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn., 2013, pp. 1–3.

[2] O.S.EyobuandD.Han, "Feature Representation Data Augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network," Sensors, vol. 18, no. 9, p. 2892, Aug. 2018.

[3] D.ThakurandS.Biswas, "Smartphone Based Human Activity Monitoring and recognition using ML and DL: A comprehensive survey," J. Ambient Intell. Humanized Comput., vol. 11, no. 11, pp. 5433–5444, Mar. 2020.

[4] R.-A.Voicu,C.Dobre,L.Bajenaru,andR.-I.Ciobanu, "Humanphysical activity recognition using smartphone sensors," Sensors, vol. 19, no. 3, p. 458, 2019.

[5] D.Anguita,A.Ghio,L.Oneto,X.Parra,andJ.L.Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in Proc. Int. Workshop Ambient Assist. Living, 2012, pp. 216–223.

[6] P.Zappi,C.Lombriser,T.Stiefmeier,E.Farella,D.Roggen,L.Benini, and G. Tröster, "Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection," in Wireless Sen- sor Networks, R. Verdone, Ed. Berlin, Germany: Springer, 2008, pp. 17–33.

[7] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. R. Del Millán, and D. Roggen, "The opportunity challenge: A bench-mark database for on-body sensorbased activity recognition," Pattern Recognit. Lett., vol. 34, no. 15, pp. 2033–2042, Jan. 2009.

[8] J.W.Lockhart,G.M.Weiss,J.C.Xue,S.T.Gallagher,A.B.Grosner,and T. T. Pulickal, "Design considerations for the WISDM smart phone-based sensor mining architecture," in Proc. 5th Int. Workshop Knowl. Discovery Sensor Data, New York, NY, USA, 2011, pp. 25–33.

[9] C. Ronao and S.-B. Cho, "Deep convolutional neural networks for human activity recognition with smartphone sensors," Lect. Notes Comput. Sci., vol. 9492, pp. 46–53, Nov. 2015.

[10] D.Anguita,A.Ghio,L.Oneto,X.Parra,andJ.L.Reyes-Ortiz, "Public domain dataset for human activity recognition using smartphones," in Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn., 2013, pp. 1–3.

[11] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "A deep learning approach to on-node sensor data analytics for mobile or wearable devices," IEEE J. Biomed. Health Inform., vol. 21, no. 1, pp. 56–64, Jan. 2017.

[12] J.R.Kwapisz,G.M.Weiss,andS.A.Moore,"Activity Recognition Using cell phone accelerometers," ACM SIGKDD Explor. Newslett., vol. 12, no. 2, pp. 74–82, May 2011.

[13] M.Bachlin,M.Plotnik,D.Roggen,I.Maidan,J.M.Hausdorff,N.Giladi, and G. Troster, "Wearable assistant for Parkinson's disease patients with the freezing of gait symptom," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 2, pp. 436–446, Nov. 2010.

[14] M.Bachlin,M.Plotnik,D.Roggen,I.Maidan,J.M.Hausdorff,N.Giladi, and G. Troster, "Wearable assistant for Parkinson's disease patients with the freezing of gait symptom," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 2, pp. 436–446, Nov. 2010.

[15] . I. Andrey, "Real-time human activity recognition from accelerometer data using convolutional neural networks," Appl. Soft Comput., vol. 62, pp. 915–922, Jan. 2017.

[16] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," Mobile Netw. Appl., vol. 25, no. 2, pp. 743–755, Apr. 2020.

[17] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in Proc. 16th Int. Symp. Wearable Comput., Jun. 2012, pp. 108–109.