International Journal of Communication Networks and Information Security
ISSN: 2073-607X, 2076-0930
Volume 14 Issue 3 Year 2022 Page 303:312

Radar Based Activity Recognition using CNN-LSTM Network Architecture

Dr. A. Helen Victoria
Assistant Professor, Department of Networking and Communications, SRM Institute of Science and Technology, Chennai, India
helenvia@srmist.edu.in
S.V. Manikanthan
Melange Academic Research Associates, Puducherry, India
prof.manikanthan@gmail.com
Dr.Varadaraju H R
Professor and HOD, Department of Electronics and Communication Engineering, Akshaya Institute of Technology, TUMKUR
hrvrsiet@gmail.com
Muhammad Alkirom Wildan
Department of Management, Faculty of Economics and Business, University of Trunojoyo Madura, Bangkalan, 69162 Jawa Timur, Indonesia
wildan.alkirom69@trunojoyo.ac.id
Kakarla Hari Kishore
Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
kakarla.harikishore@kluniversity.in

Article History

Received: 13 July 2022
Revised: 20 September 2022
Accepted: 26 October 2022

Abstract

Human Activity Recognition based research has got intensified based on the evolving demand of smart systems. There has been already a lot of wearables, digital smart sensors deployed to classify various activities. Radar sensor-based Activity recognition has been an active research area during recent times. In order to classify the radar micro doppler signature images we have proposed a approach using Convolutional Neural Network-Long Short Term Memory (CNN-LSTM). Convolutional Layer is used to update the filter values to learn the features of the radar images. LSTM Layer enhances the temporal information besides the features obtained through Convolutional Neural Network. We have used a dataset published by University of Glasgow that captures six activities for 56 subjects under different ages, which is a first of its kind dataset unlike the signals captured under controlled lab environment. Our Model has achieved 96.8% for the training data and 93.5% for the testing data. The proposed work has outperformed the existing traditional deep learning Architectures.

Keywords: Long Short-Term Memory, Radar, Micro doppler Signatures, Deep Learning

1. Introduction

Human Activity Recognition been in research for past few years. The need for recognizing daily life activities like Walking, standing, sitting, picking objects, doing a specific action to critical life-threatening activities like falling, fainting, feeling physically discomfort can be useful for varied applications right from elderly home, rehabilitation centres to defense, surveillance Monitoring. Our still existing COVID-19 pandemic has emphasized the need for Activity Monitoring due to the death rates observed in elderly homes [1]. The main reason for elderly home care death rate is due to the people who were left alone without any proper smart monitoring systems. Existing smart monitoring systems like high end smart cameras, wearable devices, and latest sensing devices [2]. But all these devices are capturing the people's images violating the privacy of the people. Another issue with camera-based sensing devices is that it might not be so effective under different environmental conditions. In order to overcome these problems, radar sensing technology has been an effective solution. Radar has unique capabilities to sense even in adverse environmental conditions like fog, dark places [3]. It is even capable of sensing through the walls. In addition, it is not a wearable device as well and it doesn't produce original images it gives signal as output. This ensures that the person's privacy is not as open as original photographs or videos of the person under study. Radar signals are converted to micro doppler signatures [4]. Micro doppler signatures are micro movements exhibited as pre-processed spectrograms that helps the obtained back scattered signals from radar to be visualized as images for the researchers to analyse. In this regard, this can be formulated as an Image classification problem. Computer vision-based image classification tasks has been handled well using Deep Learning. This is evident from the ImageNet challenge that showed excellent results using deep learning for Image Classification tasks [5]. Machine Learning Algorithms has been used by most of the previous work to classify radar micro doppler signatures [6]. Feature extraction is done by identifying the speed of walking, leg velocity prior. These features are given as input for the Machine Learning to classify the Activities [7]. Seven Activities of 12 human subjects has been classified from six features achieving an accuracy of 90%. The extracted features are handled using SVM classifier [8]. Mel frequency cepstral coefficients were used as features trained in KNN and SVM classifiers [9]. Yet the main disadvantage of Machine Learning Algorithms is that the model needs manual intervention for extracting the features for training. This requires a person to have expertise in radar domain. These issues have been sorted out by Deep Learning Algorithms itself. It learns the features through back propagation learning. Convolutional Neural Networks has surpassed SVM by 13% for classifying aided and unaided gait patterns of subjects under study [10]. This indicates deep learning algorithms has fared well than Machine Learning Algorithms with respect to Image classification in specific to radar based micro doppler image signatures. Deep Learning Algorithms like CNN, Auto encoders, pretrained networks have been used for Activity Classification. Many algorithms have obtained good accuracy but the previous results have been achieved for limited number of people maximum to 6 to 7 people. Most of the previous works have used only spatial features to classify the radar signal-based activities. Temporal information will give good insights about the activity pattern. Most of the state of art works have used CNN and LSTM separately. In this work, we have used a dataset published by University of Glasgow. The data is obtained across different people under different age groups unlike previous data collected only in the Lab Environment [11]. This is a first of its kind radar dataset that has acquired the radar signals of over 56 people. Hence working in this dataset is a challenging task as it adheres to real life scenario. The main aim of our work is to classify the activities at most accuracy. We have proposed CNN-LSTM Model Architecture to train and test the data. In addition to spatial features time-based movement variation insights are also given by LSTM layer to the model. We have also achieved a good learning in the training data. The accuracy of the testing data shows that it generalizes well with the new unseen data.

2. LSTM Networks

The main Long short-term Layer Network is to retain the behavioural history of the data under study. By default, it retains all the information but it has additional gates to forward or forget the required or necessary information. The basic work of LSTM layers is given in the Figure.1.

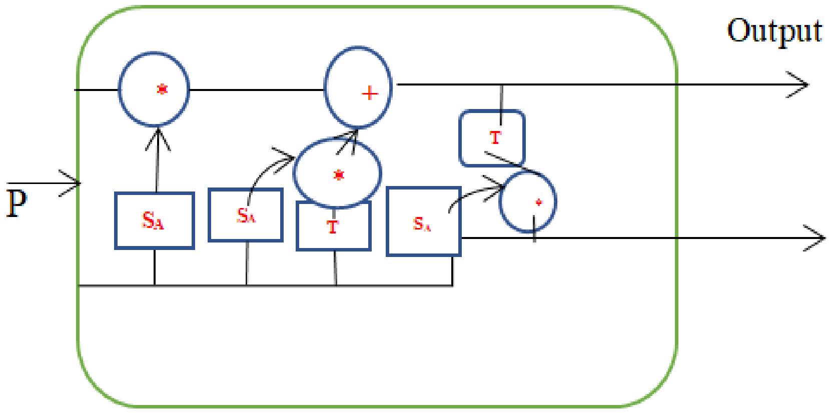

Figure 1. LSTM Internal Structure

The LSTM Network basically has 3 gates for including or discarding the neuron information to the new cell state. The first one is the forget gate which adds or eliminates information to the cell state of the Network. This gate has sigmoid activation that accepts neurons between 0 and 1 and completely fires out the neurons below 0. The Computation in the forget gate is given by the Equation.1, in which S_A represents the sigmoid Activation, W and Bias indicates the weights and bias for each input at time step X(t) and H(t-1) indicates the previous time step output.

$$fg(t) = S_A(W.[H_{(t-1)}, X(t)] + Bias) \qquad (1)$$

The second input layer shows which values has to be updated from the previous information to the cell state. This layer also has sigmoid Activation. Besides this, Tanh Activation function is used to add the newly identified candidate values to the cell state. The computations of these layers are given in the Equations [2-3].

$$I_t = S_A(W.[H_{(t-1)}, X(t)] + Bias \qquad (2)$$

$$Ca(t) = \tanh(W_c.[H(t-1), X(t)] + Bias_c) \qquad (3)$$

In Equation, 2, I_t represents the input layer output at time step t with Sigmoid Activation S_A.C_a(t) indicates the candidate values at time step t with tanh as the activation function ranging from -1 to 1. W_c and $Bias_c$ represents the Candidate weights and candidate bias respectively.

The final cell state updations are done by multiplying the old state with forget layer output, then by adding the new candidate values obtained by the Equations [2-3]. The sigmoid activation is used to find the filtered values. Tanh Activation is mainly for pushing the filtered version of the previous cell states as new candidate values. This final computation of the LSTM Layer is given by the Equation.4.

$$Nc(t) = fg(t) * C_{(t-1)} + I_t * Ca(t) \qquad (4)$$

Nc(t) represents the final output cell state value obtained by multiplying forget gate output, fg(t) and adding the candidate values output, Ca(t). Thus the LSTM layers learns the long and short term information apart from the feature extraction done by CNN to yield accurate classification results.

3. Radar Signal Representation

The radar signals are obtained as 1-dimesional complex array. The data is converted to Range bins in form of range time plot to find the pattern of transitions at each snapshot. Fast Fourier transform is used to convert the time domain of a signal to a frequency domain to obtain doppler values. Short time Fourier transform is applied to obtain clear snapshots of frequency to generate range doppler images which is called as micro doppler signature images. These spectrograms are visual representations of signal frequency spectrum variations with respect to time. Hence the Radar signals values are converted as spectrogram images which is given as input to the Deep Learning Model. The pre-processing steps are given in the Figure.2.

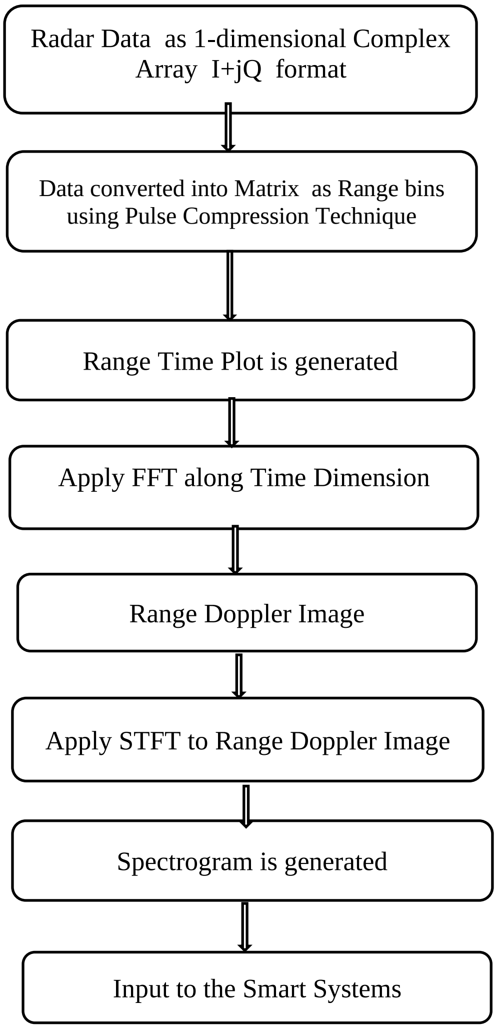
Radar Data as 1-dimensional Complex Array I+jQ format
Data converted into Matrix as Range bins using Pulse Compression Technique
Range Time Plot is generated
Apply FFT along Time Dimension
Range Doppler Image
Apply STFT to Range Doppler Image
Spectrogram is generated
Input to the Smart Systems

Figure 2. Radar Signal Pre-processing Steps

4. CNN-LSTM Network Architecture

The proposed CNN-LSTM Architecture given in the Figure.3 has two convolutional Blocks followed by the LSTM Layer and Dense Layers. The main aim of this architecture is to find the spatial as well as the temporal patterns of activities under study. The size of the input images is 64*64*3, where 64*64 represents the size of the image and 3 represents the depth of the image indicating the Red Green Blue color channels of the image. The convolutional blocks acts as a feature extractor that learns the features by itself through back propagation learning. Each convolutional block has a convolutional layer followed by a Pooling Layer. The first convolutional layers has 32 filters yielding 1568 parameters and the second convolutional layer has 16 filters yielding 8208 parameters. These are the features through which the model learns the spatial relationships among the input images. The pooling used in this work is Max Pooling which reduces the feature maps into half the count. Pooling layer is mainly used to reduce the dimension of the data and thereby filtering the most unique features to the next layer as input.

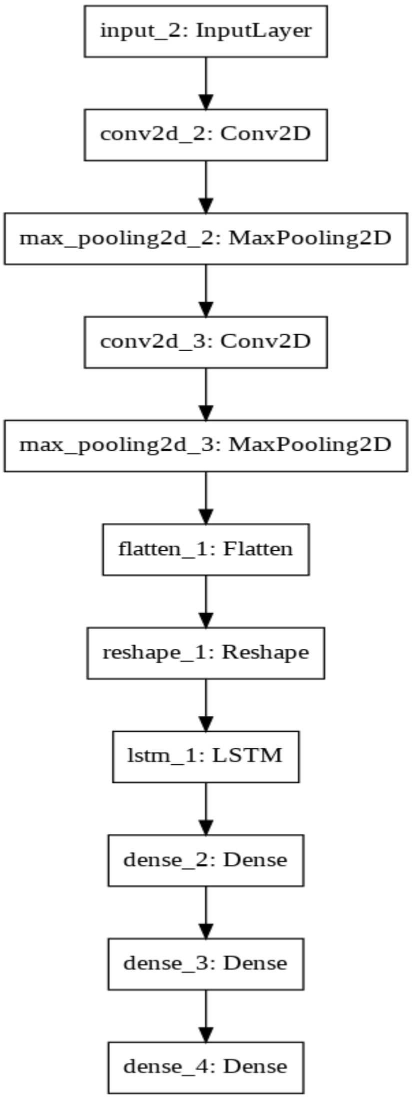

Figure 3. Proposed CNN-LSTM Architecture

The features are flattened in order to pass on to the next input layer. The features are reshaped to use LSTM Layer for predicting the sequences based on time steps. Here the spatial information is combined with the sequence-by-sequence temporal information. The LSTM Layer has 512 units and uses dropout and recurrent dropouts for each time sequence at the rate 0.3. This means 3% of neurons are fired Overfitting may yield more training classification accuracy but less testing accuracy. Inclusion of dropout in the model assures that the model is not overfitting to the training data. Hence the model achieves both the spatial and temporal insights about the data under training. The LSTM Layer is followed by three Dense Layers for one-to-one mapping of the neurons filtered. Each dense layer has different Activation Functions like RELU, Sigmoid and Softmax at the last layer for the classification of 6 classes. The Pooling, Flatten and Reshape acts as dimension reduction layer, pre-processing layers respectively. They don’t have any trainable parameters. The summary of the Architecture with input shape and its parameters at each Layer is given in the Table.1.
The total number of trainable parameters obtained through this model is 6,67,2502**.** A total of 3376 micro doppler images were used, out of which 2700 are the training images and 676, the testing images.

Table 1. Proposed Model Summary

Layer Type	Output Shape	No of Parameters	Activation Function
Input	64*64*3	0	-
Convolutional CV1	61*61*32	1568	RELU
Max Pooling 1	30*30*32	0	-
Convolutional CV2	27*27*16	8208	RELU
Max Pooling 2	13*13*16	0	-
Flatten	(None,2704)	0	-
Reshape	(None,1,2704)	0	-
LSTM	(None,512)	6588416	-
Dense D1	(None,128)	65664	RELU
Dense D2	(None, 64)	8256	Sigmoid
Dense D3	(None,6)	390	Soft Max

The Hyperparameters plays a major role in order to tune the model for faster convergence and better learning during training. In this work, Adam optimizer is used since it has minor weight updating which indicates that it doesn’t miss any important information. The optimizer shows faster learning with the convergence at 20th epoch achieving global minimum. The batch size is kept as 32 based on previous state of art approaches. The Activation Functions used are RELU in the two convolutional Layers and in the dense layer which is right after the LSTM Layer.

The Activation function RELU is given in the Equation.5.y represents neuron values greater than 0. The RELU Activation function imbibes non linearity into the training data. This makes better spatial information mapping, since the images which are given as inputs in our model is basically nonlinear data. This also helps in generalizing the data well with respect to the testing data.

$$R(y) = \max\{0, y\} \qquad (5)$$

The penultimate dense layer has Sigmoid as the Activation function. After RELU in the first dense layer sigmoid extends the activation of neurons range from -1 to +1. This yields more one to one mapping than the RELU Activation function. The Last Dense layer has softmax Activation function which helps to identify the class with the highest confidence score. This is done by taking exponentiation of each individual value and dividing it with the total sum. This is given in the Equation.6.

$$S(x_k) = \frac{e^{x_k}}{\sum_k e^{x_k}} \qquad (6)$$

In this we have 6 units in the last layer, since we have 6 classes to be classified. The class with the highest probability will be chosen as the output during training and if it does not match with the correct label given for training, the model trains to achieve the correct result through back propagation learning. This is obtained by using the Loss Function. Binary cross Entropy is used in this model. The main aim of using Binary cross entropy classification is that, it doesn’t influence one class with another class decision**.** The binary cross entropy loss, formulated is given by the Equation.7.

$$B_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i.\log(p(y_i)) + (1-y_i).\log(1-p(y_i)) \qquad (7)$$

p(y) is the predicted label. This will be applicable for each class wherein for each class all other classes will be taken as a single probability. This gives higher results than the traditional categorical accuracy. Regularization is obtained by using dropout as discussed to reduce overfitting. Parameters are initialized randomly and these hyper parameters helps to update the parameters for proper fine tuning of the model. Based on the loss function calculation, the parameters are updated using back propagation.

The filters are updated in the convolutional layers during back propagation. These filters updates makes the convolutional layer to learn the spatial features of the input images. In the other layers, the weights and bias values are updated during back propagation training. Default value of bias is retained in this model configuration. Weight initialization is retained as Xavier initialization for initial weights. We have used Google Colab GPU based run time environment. We have trained and tested the network using python Keras framework. Hyperparameter choice is made after running the model with different optimizers, Learning Rates and different Activation Functions.

5. Experimental Results

The observed spectrograms shown in the Figure.4 indicates that there is small variations among each and every activity. The Walking pattern has more oscillations from the hands and the legs than the torso movements.

Most of the Activities has similar Micro doppler image signatures. Hence there is only a minimal difference among the activities. The main challenge of the classifier is to identify these minimal differences and classify it correctly.

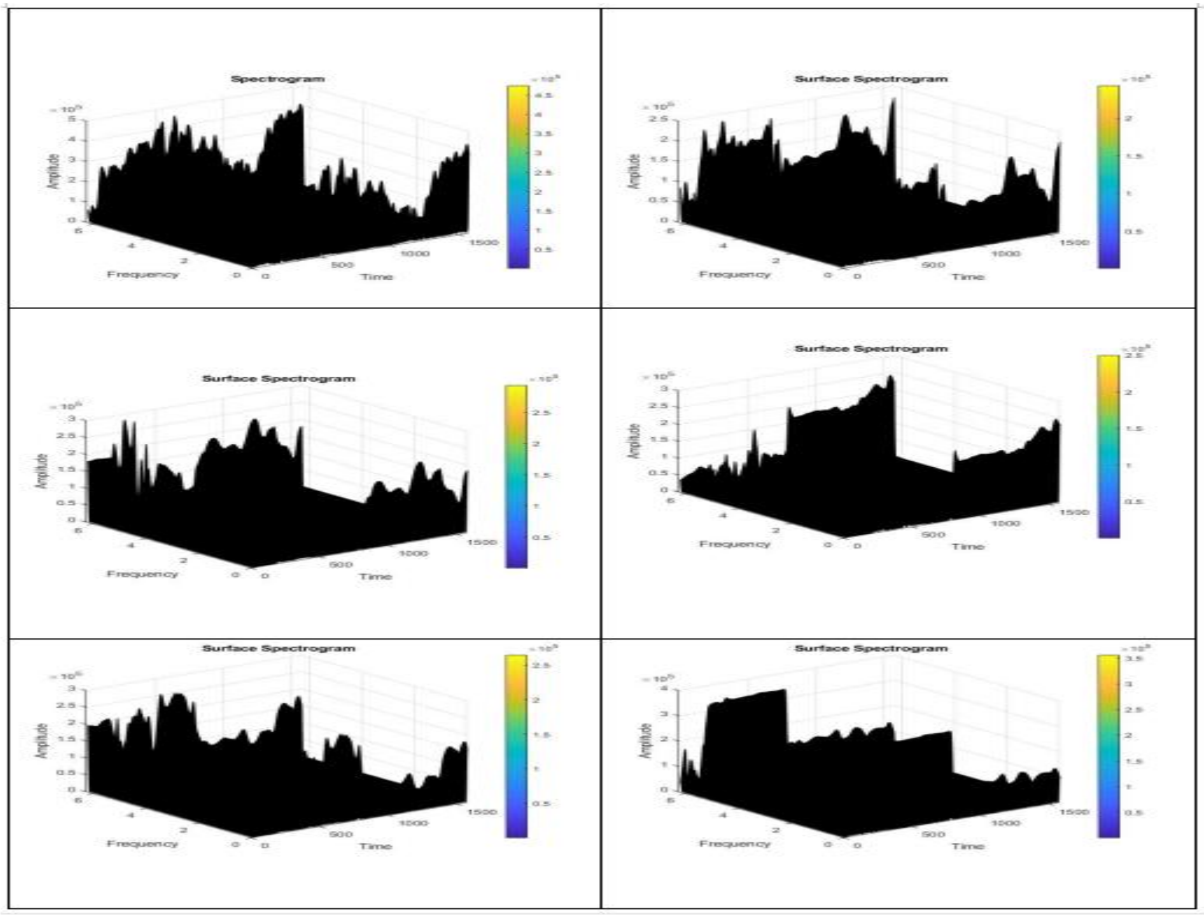

Figure 4. Surface Spectrograms of Various Activities

The feature maps are visual representations of how well our model is learning the features at each layer. Convolutional layer plays a major role in extracting the features at its each layer. Deeper the layer more unique features for a particular image is extracted. Each and every surface spectrogram indicates that there are large to minimal frequency variations, hence apart from identifying the spatial relationship temporal information through the long short term layer adds more insights to the model under training.

The Feature Maps are given in the Figure.5 represents the first convolutional, its pooling layer and the second convolutional layer and its pooling layer respectively. These feature maps indicates the features obtained and fed into the model for identifying the temporal information using LSTM Layer. This layer has input gate, output gate and forget gate. The latter gate helps in remembering only the most important information and leaving out the irrelevant information. The prior input is compared with the current output and the neurons in this output are activated using Tanh activation function. Only those activated neurons are passed to the next layer through the output cell state. This adds more uniqueness to each and every activity. This mainly alleviates the vanishing gradient problem of Convolutional Neural Networks, where the gradient becomes zero during back propagation learning. On observing the

Figure.4. It emphasizes the need for each and every time step monitoring as there are differences in movements across time for identifying the action performed by the subject under study.

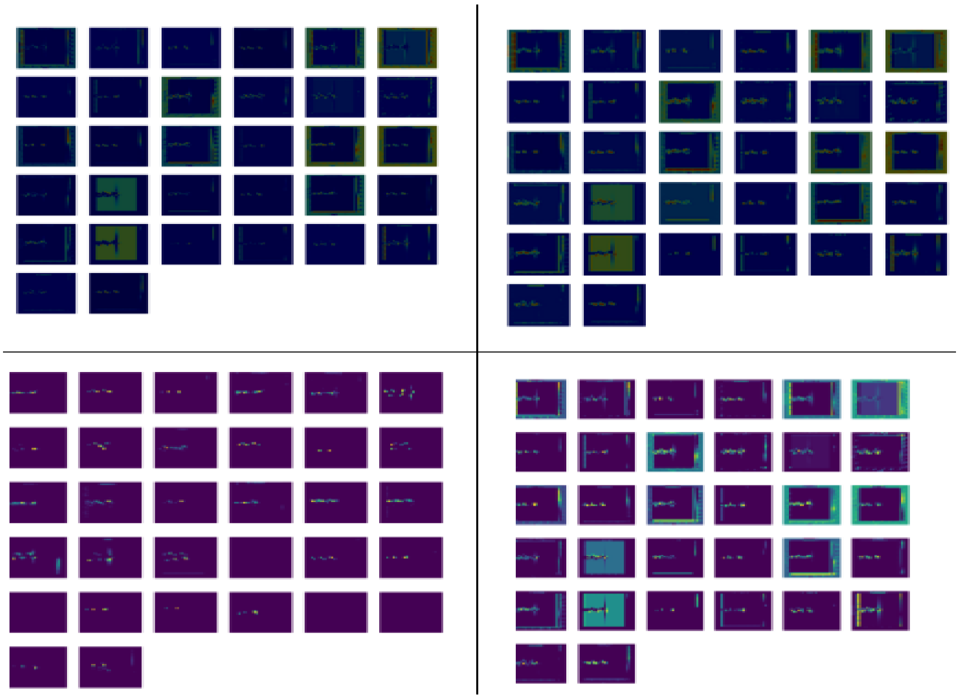

Figure 5. Feature Maps of the Proposed Model

The model has achieved an accuracy of 96.8% for the training data and 95.3% for the testing data. To eliminate overfitting dropouts and recurrent dropouts has been used in the LSTM layer. The proposed model has obtained good results with respect training data indicating that there has been proper learning in the model. The results with respect to test data indicates the model is capable of generalizing well to the newly seen test data. Convergence has been achieved in 17th epoch itself. The Accuracy plot is given in th**e** Figure.6.

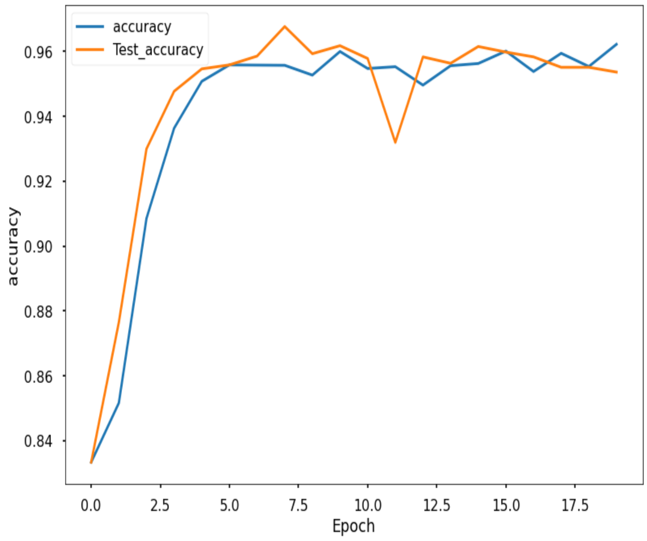

Figure 6. Training Accuracy vs Testing Accuracy of the proposed Model

Proper learning is evident from the decrease of loss at each epoch of the training and the testing data. There is a huge improvement in learning leading to decrease in loss right from the 6th epoch. The model has to be trained until there is no more learning or the learning has reached a saturation point with the training data. The loss with respect to training and the testing data is given in the Figure.7.

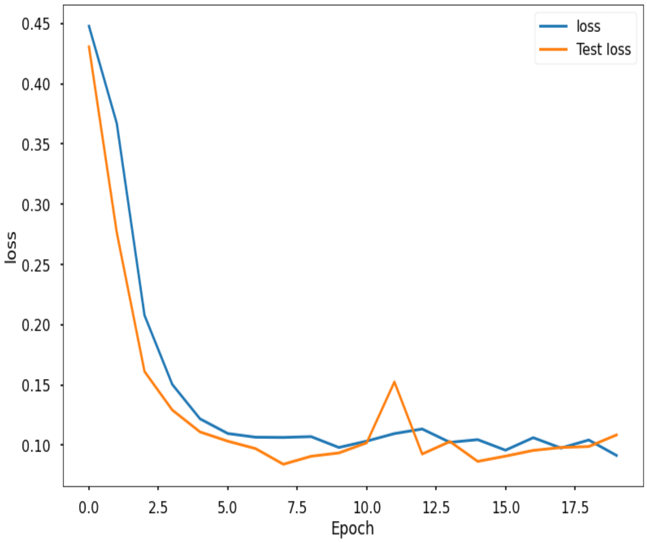

Figure 7. Training Loss vs Testing Loss of the Model

Each activity like walking, sitting and standing exhibit 100% correct classification. Taking object and drinking Water, Falling shows 95%, 93% and 98% respectively in the testing data. There are some misinterpretations among picking an object and drinking water which exhibits similar patterns. Both are indeed more or less similar and comes under common category of Activities. Falling has 2% misinterpretation in the test data, which needs further understanding of the model to be learned, since falling is the critical activity to be monitored and has to be taken care of. But in this work, accuracy has improved when compared to traditional convolutional neural network. The CNN Model achieved 91% accuracy which is 4.3% less accuracy than the test data. This improvement is mainly due to additional insights of temporal information along with the spatial information retrieved from Convolutional Neural Network.

6. Conclusion and Future Work

The different activities like walking, standing, sitting, picking an object, drinking water and falling are classified using proposed CNN- LSTM Model. In this work, by combining both CNN and LSTM temporal information along with the spatial based feature extraction adds more learning to the model. Since most of the activities has acquired good learning with prior state information, this has helped in improving the accuracy compared to the traditional CNN. Falling is a critical activity which can’t be misclassified hence there is always a need for improvising the understanding of the model. The model has achieved 95.3% accuracy with 2% mishap in the critical activity like falling. This seems the model has performed reasonably well with respect to training and the testing data. This can be deployed in real time applications like elderly home care, hospital monitoring and small office groups.

References

[1] Md. publishes data on COVID-19 cases at nursing homes, group facilities, (28 April 2020).https://wtop.com/maryland/2020/04/md-publishes-data-on-covid-19-cases-at-nursing-homes-group-facilities/.

[2] Yang.X.; Tian.Y. Super Normal Vector for Human Activity Recognition with Depth Cameras(2017).IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 5, pp. 1028-1039, doi:10.1109/TPAMI.2016.2565479.

[3] M. Amin, Moeness G(2017).Radar for Indoor Monitoring: Detection , Classification, and Assessment. 1st Edition, Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781315155340.

[4] Chen, V.C. Radar Micro-Doppler Signatures: Processing and Applications, Electromagnetics and Radar Book series.(2014).Radar, Sonar & Navigation. Institution of Engineering and Technology, ISBN-101849197164.

[5] Jindong Wang.; Yiqiang Chen.; Shuji Hao.; Xiaohui Peng.; Lisha Hua.(2019).Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, Volume 119, Pages 3-11, https://doi.org/10.1016/j.patrec.2018.02.010

[6] Qingchen Zhang Laurence T.; YangZhikui ChenPeng Li(2018).A survey on deep learning for big data, Information Fusion, Volume 42, pages 146-157. doi.org/10.1016/j.inffus.2017.10.006.

[7] Saho, K.; Uemura, K.; Sugano.K.; Matsumoto, M(2019). Using Micro-Doppler Radar to Measure Gait Features Associated With Cognitive Functions in Elderly Adults. IEEE Access, vol. 7, pp. 24122-24131, doi: 10.1109/ACCESS.2019.2900303.

[8] Youngwook Kim and Hao Ling(2009). Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine.IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING,VOL. 47, NO. 5, DOI: 10.1109/TGRS.2009.2012849.

[9] Liang, L, Mihail, P,Marylin, R, Marjorie, S,Paul(2011). Tarik, Y. Automatic fall detection based on doppler radar motion signature.5th International Conference on Pervasive Computing Technologies for Healthcare, pp. 222-225, doi: 10.4108/icst.pervasivehealth.2011.245993.

[10] Goodfellow, I. Bengio, Y, Courville, A and Bach, F(2016). Deep Learning. Adaptive Computation and Machine Learning series .Cambridge,USA; MA: MIT Press.

[11] Fioranelli, F. , Shah, S. A. , Li, H., Shrestha, A., Yang, S. and Le Kernec, J. (2019). Radar signatures of human activities,University of Glasgow,DOI:10.5525/gla.researchdata.848. http://researchdata.gla.ac.uk/848/.