ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Recognizing Human Activity Using Hybrid Models of CNN and LSTM in Deep Learning

Dr. G. Krishna Mohan[1], Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
N. Gowthami[2], M. Lakshmi Tulasi[3], M. Geethika[4], P. Komala Jyothi[5]
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
gvlkm2002@gmail.com[1], gowthamireddy091@gmail.com[2],
paiditalli.tulasi@gmail.com[3], geethika.mundlamuri1@gmail.com[4],
may28komalajyothi@gmail.com[5]

DOI:10.48047/IJFANS/V11/I12/178

Abstract

Human Activity Recognition (HAR) is a time series categorization challenge that requires data from a number of timesteps in order to correctly classify the activities that are carried out. In recent times, the usage of image datasets for activity recognition has increased, however good classification cannot be done with just one frame. To increase recognition accuracy, multiple frames of data and the context of environment are required. It is known that a video is made up of a number of still images (frames) that are quickly updated to create the illusion of motion. The hybrid models of Deep Learning (DL) algorithms like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) are proposed for recognising the human activity from video dataset. The hybrid models, Convolutional Long Short-Term Memory (ConvLSTM) and Long-term Recurrent Convolutional Network (LRCN) are introduced to improve the accuracy of HAR on video dataset. The models will be evaluated on standard video datasets, and the outcomes will show how HAR has the potential to significantly influence a number of industries. It has many applications in the fields of security, sports, and healthcare.

Keywords: Convolutional Neural Networks (CNN), Convolutional Long Short-Term Memory (ConvLSTM), Human Activity Recognition (HAR), Long Short-Term Memory (LSTM), Long-term Recurrent Convolutional Network (LRCN).

1. Introduction

HAR from video data has received a lot of attention in recent years, thanks to the availability of large-scale video datasets and the advancement of deep learning techniques. Video-based HAR has several advantages over image datasets and sensor modalities like accelerometer and gyroscope because it can capture a more complete view of human activities, including temporal and spatial information. The context and environmental signals that can help in accurately identifying human actions can also be captured in video footage. Deep learning models have demonstrated notable performance across a range of computer vision applications, including instance segmentation, semantic segmentation,

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

object detection, and image categorization. Deep learning models have also been used to HAR tasks, and recent research has demonstrated that these models can perform well on a variety of HAR datasets. Yet due to a number of issues, those models are not very effective. Then hybrid deep learning models began to develop to address those issues. A comparison of hybrid deep learning models is done for recognising human activities using video datasets in this research. Assessing the effectiveness of various deep learning models using the well-known UCF-50 video dataset. Also, we assess the accuracy and confidence of both models while testing them on dynamic videos to determine which one is the best.

2. Literature Survey

The development of sensor data and image recognition algorithms, which have been repeatedly altered and expanded to cope with video data, has been a primary driving element in video recognition research. The authors considered a waist wearable device to collect data [1]. Standard wireless sensor data mining dataset (WISDM) is used to recognize human activity [2]. The main vision of action recognition was to use in security surveillance it detects with the normal actions and the abnormal actions [3-4]. The accelerometer and gyrometer data were also considered to recognize human activity [5-6]. Initially for video dataset single stationery camera was used to capture activities [7].

It has been noticed that the actions with low variations in background were considered as the dataset [8]. Deep learning models like 1D CNN and 1D LRCN were tested on WISDM dataset which have 6 regular activities [9]. Neural network models were also applied for HAR tasks, performance of ANNs is greatly influenced by their hyperparameters and input data [10-11]. A lot of comparative and exhaustive analysis were made on the dataset to find factors that effects the recognition [12]. Machine learning algorithms have shown good accuracy for recognition on different types of datasets [13-15].

It is considered that Deep learning models, have tested for recognition of human activity on both sensor dataset and video dataset, even hybrid models such as LSTM-RNN were considered [16-18]. Lots of video dataset in prior were staged actors, but recently HAR tasks focused on the realistic videos even [19]. This article uses the unsupervised learning technique, on the dataset which includes both spatial and temporal features in them, have shown better results than the supervised learning technique [20].

It was mentioned that for input datasets even flow of histograms were considered in order to perform the HAR tasks [21]. Considering removing noise as one of the major pre-processing techniques and have removed the irregularities in the sequence of images of a

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

video [22]. It is observed that temporal templates have played a key role in detection of human activity [23]. The Hybrid LSTM network was considered for HAR tasks based on the smart watch dataset [24].

It was mentioned that LSTM was typically built in "blocks" or "cells," each of which has three or four gates, such as a forget gate, an input gate and an output gate. In order to solve the vanishing gradient problem, LSTM employs the idea of gating. The cell can remember values over a variety of time intervals [25], [26-34].

3. Problem identification

HAR using video datasets needs correct classification of human activities, but current approaches using a random snapshot from a video or an image can lead to incorrect classifications due to the complexity of actions and lack of contextual information. Deep learning models must be created using video data to extract time and space-related information to improve accuracy. By creating hybrid models, ConvLSTM and LRCN for reliable HAR utilising video datasets, this effort seeks to address this issue.

4. Proposed Methodology

CNNs are wonderful with image data while LSTMs are fantastic with sequence data. Combining the strengths of CNNs with LSTMs produces efficient video categorization for difficult computer vision problems. An LSTM model is used to determine the temporal relationships between frames after a CNN model extract features from the video. The two hybrid models of CNN and LSTM are ConvLSTM and LRCN.

4.1 Data Collection

The UCF50-Action Recognition dataset is considered and it comprises of 50 different action categories with a total of 6,618 video files in it. In this case, the dataset is divided 75% to 25% into training set and testing set. However, due to computational and size constraints, we can only consider a set of action categories.

4.2 Data Resizing and Normalizing

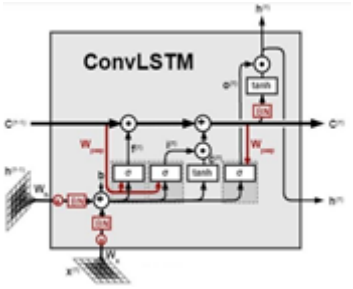
Resizing and normalising are two methods used for pre-processing the dataset in our proposed models. Resizing is the process of adjusting a frame size to a desired size, usually to make it compatible with the input size of a model. Normalisation scales a frame's pixel values to a predetermined range, enhancing a model's performance by ensuring input data consistency. This technique reduces CNN training time and improves accuracy.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

4.3 ConvLSTM model

An LSTM network variation that incorporates convolutional operations is known as a ConvLSTM cell. It is an LSTM with embedded convolution, which enables it to recognise spatial aspects of the data and considering the time relationship. This method efficiently captures both the spatial and temporal relationships present inside individual video frames in order to classify videos. The ConvLSTM can model spatio-temporal data independently and can accept 3-D input because of this convolution structure.



4.4 LRCN model

In a single model, the LRCN integrates the CNN and LSTM layers. The spatial information is extracted from the frames using convolutional layers and supplied to the LSTM layers at each time step in order to simulate the temporal sequence. By explicitly learning spatiotemporal features throughout an end-to-end training, the network creates a robust model. Also, a

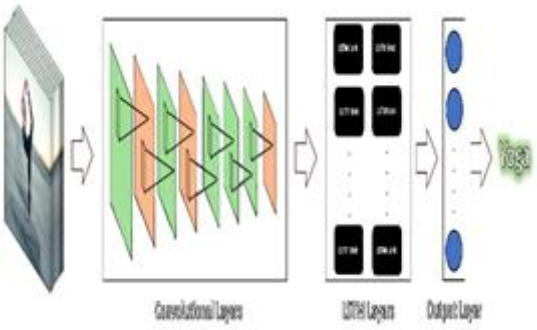


Figure 1: ConvLSTM cell

Time Distributed wrapper layer is used. It enables us to independently apply the same layer to each frame of the video. As a result, the layer can now capture the entire video into the model in a single shot.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Figure 2: working of LRCN

5. Implementation

The proposed process for implementing human activity recognition using ConvLSTM and LRCN from video dataset:

- **Data gathering**: Gather video files of different people doing different activities from UCF50 dataset.
- **Resize and Normalize**: To verify that the images have the same dimensions, resize them to a standard size, i.e. (64,64). Scalethe pixel values of the pictures to be between 0 and 1 to normalize them.
- **Splitting Dataset**: The dataset is split into two sets, a training set and a testing set. In this scenario, the dataset is split into training and testing data at a ratio of 75% to 25%.
- **Build ConvLSTM model**:Construct a ConvLSTM network to classify the human activity. The 2D Convolutional layers, ConvLSTM layers and fully connected layers all should be present in the model.
- **Build LRCN model**: Construct an LRCN network to classify the human activity. The 2D Convolutional layers, LSTM layers and fully connected layers all should be present in the model.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

- **Models Training**: Use the pre-processed dataset to train both models. Employ a suitable loss function, such as categorical cross-entropy, and an Adam optimizer.
- **Models Evaluation**: Using a different validation dataset, assess how well the trained model performed. Use metrics such as accuracy to judge the model's performance.
- **Models Testing**: Use a different testing dataset to test the final model. Verify the model's ability to generalize to the new data.

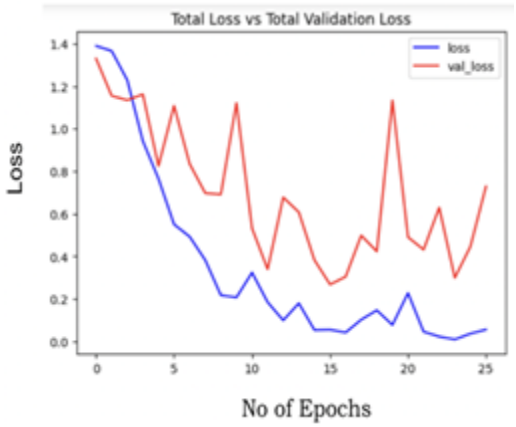
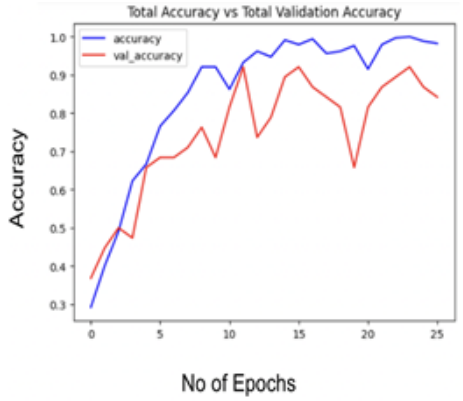


Figure 3: Flow Chart of Proposed System

6. Results & Conclusion

The activity being performed was recognised using both the ConvLSTM and LRCN models and achieved accuracies of 83.46% and 92.13% respectively. Since, LRCN model has demonstrated better performance than ConvLSTM model, LRCN is the superior model for recognizing human activities. A set of actions are only considered from UCF50 dataset and observed that the accuracies, when taken 3, 4, 5, 7, 10 action categories are 91.67%, 92.13%, 90.85%, 87.78%, 81.37% respectively on LRCN model. It is seen that the accuracy is increased from 3 to 4 and constantly decreased from 5 to 10 action categories because the learning rate is very high and considered 4 action categories as local minimum where the accuracy is maximum for the models.



-**Figure 4:** ConvLSTM Loss curves

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

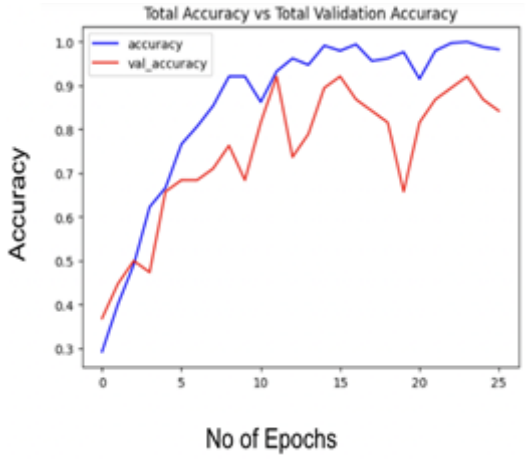


Figure 5: ConvLSTM Accuracy curves

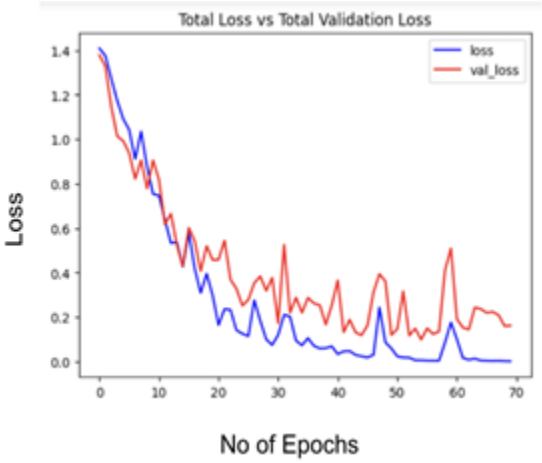


Figure 6: LRCN Loss curves

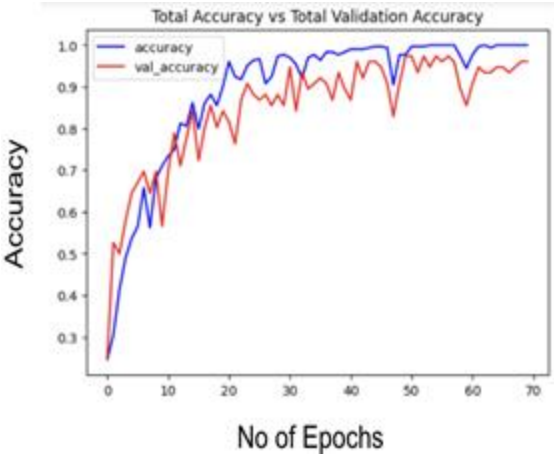


Figure 7: LRCN Accuracy curves

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

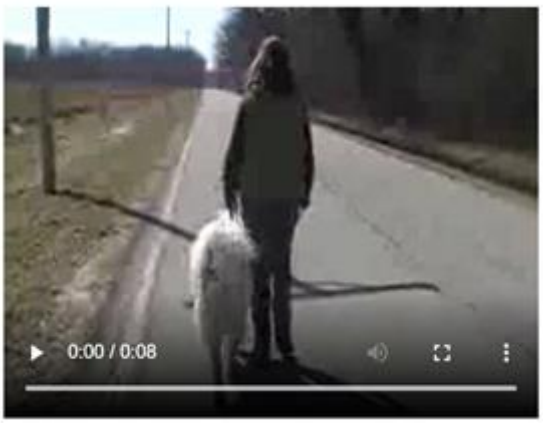


Figure 8: A video clip comprises of walking with dog activity.

A ConvLSTM model output for Figure 8:

Action Recognized: Walking with Dog

Confidence: 0.7413442134857178

An LRCN model output for Figure 8:

Action Recognized: Walking with Dog

Confidence: 0.998050332069397

The LRCN model recognizes the class with greater confidence than the ConvLSTM model. After carefully examining the data, we can conclude that the LRCN model is a better model for Human Activity Recognition.

Figure 9: A video clip comprises of Pushups activity.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Figure 9 displays the output of a video clip when the models are applied to the input video clip resulting in a labelled video as output. Here, the input video clip features Pushups, resulting in a labelled video titled as Pushups.

7. Limitations & Future Scope

These models need a lot of computes, especially when training on huge datasets. This may restrict their use in areas with limited resources. We are only able to take into account a subset of action category choices due to computational and space limitations. It is unable to categorise multiple activities simultaneously**.**

Future Scope

The potential for this idea is immense. First off, pre-processing methods like noise removal can be used to further fine-tune video recognition. To further improve the model's accuracy, far larger datasets, live videos and more realistic videos in different formats with more duration can be considered.

References

[1] Yen, C. T., Liao, J. X., & Huang, Y. K. (2020). Human daily activity recognition performed using wearable inertial sensors combined with deep learning algorithms. IEEE Access, 8, 174105-174114.

[2] Shiranthika, C., Premakumara, N., Chiu, H. L., Samani, H., Shyalika, C., & Yang, C. Y. (2020, December). Human Activity Recognition Using CNN & LSTM. In 2020 5th International Conference on Information Technology Research (ICITR) (pp. 1-6). IEEE.

[3] Zaidi, S., Jagadeesh, B., Sudheesh, K. V., & Audre, A. A. (2017, September). Video anomaly detection and classification for human activity recognition. In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) (pp. 544-548). IEEE.

[4] Zhong, H., Shi, J., & Visontai, M. (2004, June). Detecting unusual activity in video. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. (Vol. 2, pp. II-II). IEEE.

[5] 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.

[6] Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter, 12(2), 74-82.

[7] Samir, H., Abdelmunim, H., & Aly, G. M. (2017, December). Human activity recognition using shape moments and normalized fourier descriptors. In 2017 12th

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved,** UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

International Conference on Computer Engineering and Systems (ICCES) (pp. 359-364). IEEE.

[8] Hou, Y. L., & Pang, G. K. (2010). People counting and human detection in a challenging situation. IEEE transactions on systems, man, and cybernetics-part a: systems and humans, 41(1), 24-33.

[9] Cho, H., & Yoon, S. M. (2018). Divide and conquer-based 1D CNN human activity recognition using test data sharpening. Sensors, 18(4), 1055.

[10] Navneet, D. (2006). Human detection using oriented histograms of flow and appearance. Computer Vision-ECCV 2006, 3952, 428-441.

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

[12] Oniga, S., & József, S. (2015). Optimal recognition method of human activities using artificial neural networks. Measurement Science Review, 15(6), 323-327.

[13] Oh, S., Ashiquzzaman, A., Lee, D., Kim, Y., & Kim, J. (2021). Study on human activity recognition using semi-supervised active transfer learning. Sensors, 21(8), 2760.

[14] Marinho, L. B., de Souza Júnior, A. H., & Rebouças Filho, P. P. (2017). A new approach to human activity recognition using machine learning techniques. In Intelligent Systems Design and Applications: 16th International Conference on Intelligent Systems Design and Applications (ISDA 2016) held in Porto, Portugal, December 16-18, 2016 (pp. 529-538). Springer International Publishing.

[15] Hofmann, C., Patschkowski, C., Haefner, B., & Lanza, G. (2020). Machine learning based activity recognition to identify wasteful activities in production. Procedia Manufacturing, 45, 171-176.

[16] Kuppusamy, P., & Harika, C. (2019). Human action recognition using CNN and LSTM-RNN with attention model. Int. J. Innov. Technol. Explor. Eng, 8, 1639-1643.

[17] Chen, Y., Zhong, K., Zhang, J., Sun, Q., & Zhao, X. (2016, January). LSTM networks for mobile human activity recognition. In 2016 International conference on artificial intelligence: technologies and applications (pp. 50-53). Atlantis Press.

[18] Zebin, T., Sperrin, M., Peek, N., & Casson, A. J. (2018, July). Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. In 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 1-4). IEEE.

[19] Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008, June). Learning realistic human actions from movies. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved,** UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[20] Kwon, Y., Kang, K., & Bae, C. (2014). Unsupervised learning for human activity recognition using smartphone sensors. Expert Systems with Applications, 41(14), 6067-6074.

[21] Dalal, N., Triggs, B., & Schmid, C. (2006). Human detection using oriented histograms of flow and appearance. In Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II 9 (pp. 428-441). Springer Berlin Heidelberg.

[22] Boiman, O., & Irani, M. (2007). Detecting irregularities in images and in video. International journal of computer vision, 74, 17-31.

[23] Bobick, A. F., & Davis, J. W. (2001). The recognition of human movement using temporal templates. IEEE Transactions on pattern analysis and machine intelligence, 23(3), 257-267.

[24] Mekruksavanich, S., & Jitpattanakul, A. (2020, October). Smartwatch-based human activity recognition using hybrid lstm network. In 2020 IEEE SENSORS (pp. 1-4). IEEE.

[25] Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN architecture for human activity recognition. IEEE Access, 8, 56855-56866.

[26] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[27] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[28] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[29] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

[30] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[31] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[32] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[33] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[34] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.