Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 6s (2025)

Deep Custom Transfer Learning Models for Recognizing Human Activities via Video Surveillance

Saurabh Gupta[1], Dr. Rajendra Prasad Mahapatra[2]

[1,2] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Delhi NCR Campus, Ghaziabad, Uttar Pradesh

saurabhg1@srmist.edu.in[1] , rajendrm1@srmist.edu.in[2]

Article History:

Received: *16-10-2024*

Revised: *30-11-2024*

Accepted: *10-12-2024*

Abstract:

The use of video surveillance for human activity recognition (HAR) in inpatient rehabilitation, activity recognition, or mobile health monitoring has grown in popularity recently. Before using it on new users, a HAR classifier is often trained offline with known users. If the activity patterns of new users differ from those in the training data, the accuracy of this method for them can be subpar. Because of the high cost of computing and the lengthy training period for new users, it is impractical to start from scratch when building mobile applications. The 2DCNNLSTM, Transfer 2DCNNLSTM, LRCN, or Transfer LRCN were proposed in this paper as deep learning and transfer learning models for recognizing human activities via video surveillance. The Transfer LRCN scored 100 for Training Accuracy and 69.39 for Validation Accuracy, respectively. The lowest Validation Loss of 0.16 and the Lowest Training Loss of 0.001 was obtained by Transfer LRCN, respectively. The 2DCNNLSTM has a 98.34 lowest training accuracy and a 47.62 lowest validation accuracy.

Keywords Deep learning, Transfer learning, Video Surveillance, and Human Activity Recognition

1. Introduction

Because there are so many possible uses for human action recognition, academics have devoted a lot of attention to studying it during the past ten years. These include autonomous driving, surveillance, ambient-supported living, entertainment, & human-computer interaction (HCI). Two basic approaches exist for activity recognition: learning-based representations and the more traditional, manually produced feature-based representation [1–4]. The learning-based representation, as well as particularly, deep learning, which included a trainable feature extractor with a trainable classifier, introduced the idea of end-to-end learning first. The tremendous development for action recognition in videos has been disclosed by deep learning-based algorithms. For classifying images, recognizing objects, and recognizing actions, deep learning models like CNN [5] and Deep Belief Networks (DBNs) [6] were introduced to reduce the dimensionality of the input. However, building a new deep learning model from the start involves a significant quantity of data, powerful computing power, and hours, or even days, of training [7-8]. In practical applications, collecting and annotating a sizable volume of domain-specific data requires a lot of time and money. Because it might not always be possible to collect a sizable amount of domain-specific data, applying deep learning models can be challenging. To solve this issue, researchers changed the way they categorize images so that they more closely resemble how the human visual system works. Humans can learn numerous categories over their lifetimes from a small number of samples. Humans are said to have developed this skill through amassing information over time and using it to learn new things [9-10]. According to researchers, understanding earlier things helps with learning new ones because of how they are related to and similar to the new ones. Studies have shown that deep learning models which have been trained for a single classification task may be used for numerous classification challenges. CNN models that have been trained on a certain dataset or task can therefore be modified for a new task in a different domain.

This concept is referred to as domain adaptation or transfer learning. [11]Transfer learning is a machine learning technique that has long been researched for addressing various visual classification problems. The current growth of information, including images, sounds, and videos, available online has increased the demand for high accuracy and computing efficiency [12]. These reasons have contributed to the widespread use of transfer learning in the domains of machine learning and computer vision. Once conventional machine learning algorithms have reached their limits, transfer learning opens up new possibilities for visual classification. It has mostly changed the process that machines utilize to learn and handle classification tasks [13]. It has been utilized successfully for visual categorization tasks in the fields of recognizing objects, image classification, or human activity recognition. Transfer learning often employs two techniques: 1) keeping the pre-trained network in place while changing the weights in response to fresh training data. 2) using a pre-trained network for feature extraction with representation and a general classifier, like SVM, for classification [14-15]. Numerous tasks involving recognition and categorization have seen success using the second strategy. The second group also applies to our suggested method for identifying human action. We looked into recently put forward benchmark deep model examples like Alex Net and Google Net. Using the results of the studies, 2DCNN and LRCN were selected as the sources for a target model that would be used to recognize actions [16]. A 2D CNNLSTM classifier hybrid was employed for feature extraction while the representation using the source model, and the source model were subsequently used for action recognition.

2. RELATED WORK

Hejazi 2022 et al. This work seeks to recognize human behaviours by generating manually constructed features using phase data recovered from the frequency domain on the video data. We wish to explain the motion dynamics of the scene using localized phase information rather than calculating motion vectors or estimating optical flow. Phase correlation data from each of the next two frames of each of the video recordings' two consecutive frames were used to train a model to represent the action. We have worked very hard to evaluate the performance of our method on three huge and complex datasets: UCF101, Kinetics-400, and Kinetics-700. The results show that, over all of these datasets, our technique is very accurate at detecting human actions [17]. Cui 2022 et al. To lessen the pressure on the primary cloud server, the transmission load, and the processing latency of the video surveillance system, a distributed computing design is developed that directly handles peripherals' video data. A lightweight neural network model and the federated learning strategy are both advised. For a variety of circumstances, computation models are built using light neural network technology, and the produced models are logically structured in edge devices. Human motion analysis has developed into a significant research area as technology advances. This analysis makes it possible for people to recognize motion. Chinese Taekwondo is now of higher quality than it was in earlier eras due to the growth of society. Since then, the number of Chinese Taekwondo classes has gradually expanded, however, the issue with inconsistent instruction quality is to blame. The author conducted specific research using D.A. Cooper's empirical theory of learning to help with overcoming the teaching issues in the taekwondo learning process and raise the standard of Chinese taekwondo. Further investigation will be conducted to meet real-time needs, including the development of neural networks for compression acceleration algorithms, techniques for universal neural networks with weight updates, finding from detection, or federated learning algorithms. This study makes use of scientific and technological approaches to investigate technical actions and strategies. Then, these techniques are used for specific educational experimentation and assessment [18].Mar-Cupido 2022 et al. Use four pre-trained deep learning algorithms (Nas Net Mobile, MobileNetv2, ResNet101v2, & ResNet152v2) to categorize photographs according to the kind of face masks (KN95, N95, surgical, and fabric) worn by people in the pictures. Deep residual network models (ResNet101v2 and ResNet152v2) provide the

greatest performance, and the highest accuracy, with the least loss, as shown by the results of experiments [19]. Sarveshwaran 2022 et al. A person's action is intended to be recognized by human activity recognition (HAR), which is based on a set of sensor measurements. Economic and non-economic elements can both be used to classify human activity recognition. The economic kind is employed to bring in money. The non-economic type is utilized to promote mental well-being. Applications for identifying human activity include but are not limited to, smart homes, smart traffic, aged care, thief detection in public spaces, convalescence, and people work assessments. This study discusses the deep learning model, benefits, and dataset utilized for recognizing human behaviour. Using deep supervised and deep unsupervised learning models, the main HAR issues have been gathered [20]. Kumar 2022 et al. The majority of learning algorithms that have been proposed have compared activity data to the fuel used in automobiles. Having enough activity data at hand can increase the ability of learning algorithms to recognize activity patterns. This study contains data from smartphone sensors (accelerometer & gyroscope) worn around the subjects' waists during the activities, as well as the FLAAP (Finding and Learning the Associated Activity Patterns) activity dataset. In this dataset, ten behaviours were noted by eight distinct participants. A steady 100Hz sampling rate was used to gather millions of samples of raw sensor activity data between February 1 and May 31 of 2022. It was difficult to identify the Activities for Daily Living (ADL) using the HAR (Human Activity Recognition) data sets, which keep track of these behaviours and discover relationships between activity patterns. The FLAAP dataset closes this gap and can be used to identify the associated ADL patterns. The results of the experiment show that the learning technique Random Forest (RF), which was applied, correctly detected the activities with a degree of accuracy of about 77.22%. By creating more delicate learning models that will improve recognition rates, researchers can fill a research need in this area by utilizing the FLAAP dataset's applied RF learning technique. The scientific community may also take a keen interest in studying how algorithms pick up new information when employing different data pre-processing approaches, as well as knowledge transfer to target domains, and further methods [21].

3. PROPOSED METHODOLOGY

In this discussion, the methodology for human activity recognition is discussed. First, data from the UFC 50 is collected, after which 7 classes are chosen. Next, the data is pre-processed with frame capture, data resizing, and data scaling. Finally, One Hot encoding is performed, followed by modeling using 2DCNN, LSTM, and LRCNN [22-23]. Finally, the prediction is evaluated. Fig. 1 Shows the proposed flowchart and methodology and One Hot encoding is shown in equation 1.

$$\sum_{i=1}^{M}(yi - \sum_{j=0}^{p} w_j * x_{ij})^2 + \lambda \sum_{j=0}^{p} W_j^2 \quad (1)$$

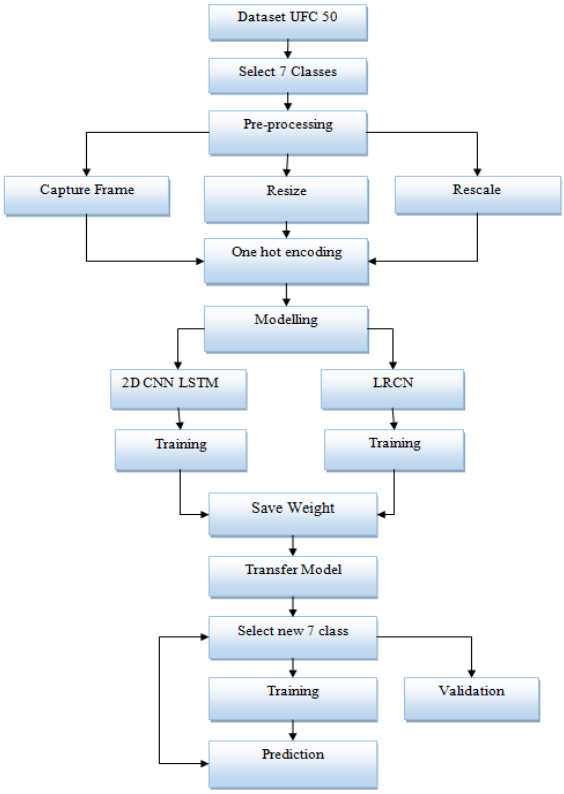


Figure 1: Proposed Flowchart

A. Data Collection

The website at [24] https://www.crcv.ucf.edu/data/UCF50.php was used to gather the information. UCF50 is a set of action recognition data that includes 50 action categories that were taken directly from real YouTube videos. This data set's development, the YouTube Action information set (UCF11), comprises eleven distinct action types. The UCF50 data collection on 50 activity categories on YouTube includes: Some examples of athletics include bench pressing, biking, shooting baskets, swimming the breaststroke, and biking. Diverting, drumming, fencing, golfing, playing the guitar, high jumping, horse racing, riding a horse, juggling balls, jumping rope, jumping jacks, kayaking, lunges, participating in a military parade, mixing batter, nun chucks, playing the piano, tossing pizza, pole vaulting, as well as pommel horse are a few of the additional activities. Punches, pullups, and pushups You can do sports like rowing, salsa spinning, skateboarding, skiing, ski jet, soccer juggling, swinging, playing the table, tai chi, tennis swinging, trampolining, playing the violin, volleyball spiking, walks with a dog and yo-yo.

B. Pre-processing

Get all the videos, capture frames from the videos, overlay text, resize frames with 64 heights as well as 64 widths, consider the length of the sequence at 20, get all classes with a list, extract frames from the data, and normalize all frames via pixel division 1255. Finally, convert all numerically categorical data to one Hot encoding.

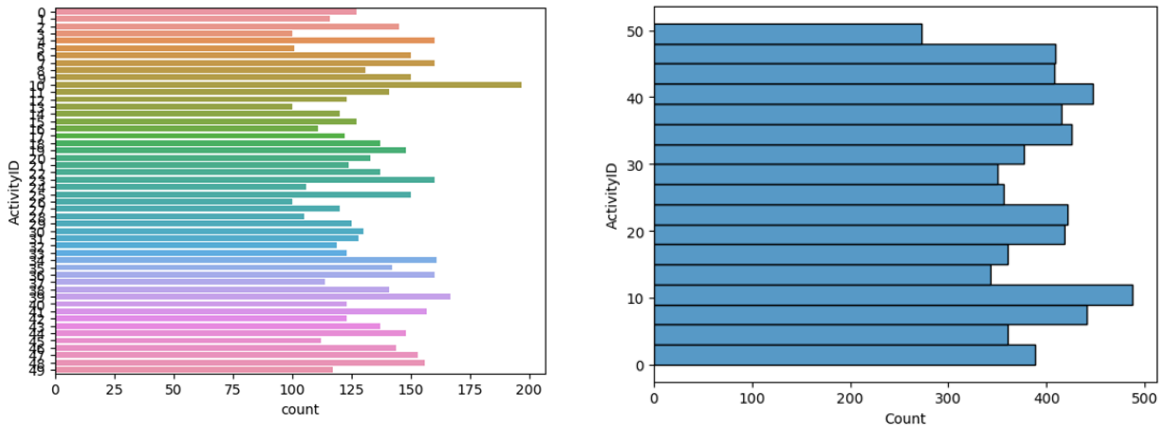
C. Data Splitting

Data have been divided 80:20. Training takes up 80% of the time, whereas testing only occupies 20%. Using the machine learning (ML) data splitting technique, overfitting can be prevented. When a machine learning system can accurately fit its training data but not any new data, this is referred to as overfitting. This circumstance falls within that group. It is typical to divide this initial data into three

to four different subgroups before putting it into an ML model. The testing and training datasets are examples of common datasets. To improve the amount of training data for each distinct data collection, it is recommended that the data be divided.

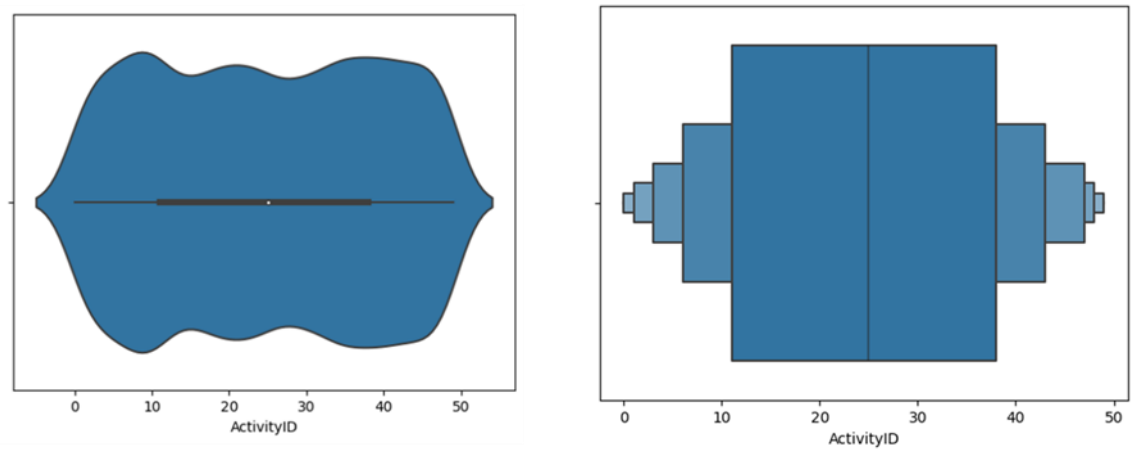
D. Exploratory Data Analysis (EDA)

Exploratory data analysis, sometimes known by its acronym EDA, is an important stage in the data analysis process. The process entails evaluating, cleansing, and displaying the data in order to find patterns, relationships, and insights in it. EDA enables analysts and data scientists to gain a better understanding of the primary attributes of a dataset and to make educated decisions regarding whether or not to proceed with additional analysis or modelling.



(a) (b)

Figure 2: Count plot (a) and histogram (b) of Activity ID



(c) (d)

Figure 3: Violin Plot (a) and Boxplot (b) of Activity ID

Fig.2 and 3 show the Countplot, histogram, violin plot, and Box plot of Activity ID. in the histogram, the greatest value exhibited is 500, which is in the range of 0 to 10, and the violin plot and box plot indicate the distribution of the values.

E. Deep learning & Modeling

Deep learning refers to the method by which a computer model directly picks up categorization skills from images, text, or sound. Deep learning models have the potential to achieve cutting-edge accuracy, occasionally surpassing human performance. Models are trained using multi-layer neural network

architectures and a massive quantity of tagged data. Implement two models in this work, such as 2DCNN LSTM & LRCN.

- **2DCNN LSTM**

Use Convolutional LSTM 2D layers alongside kernel size 3*3, filter 4, activation function "tanh," recurrent dropout 20%, input shape 64*64*3, and a time distributed dropout with 20% to create a total of 4 sections layered architecture. The output layer includes an activation function called SoftMax. The model was built with categorical cross-entropy loss, Adam was the optimizer, accuracy metrics were used, batch size was 16, validation split was 20%, and epochs were 100. Model architecture for the 2D CNN LSTM [25] is shown in Fig.4. Equation 2 displays a 2D CNN [26], while Equations 3 through 8 show LSTM equations.

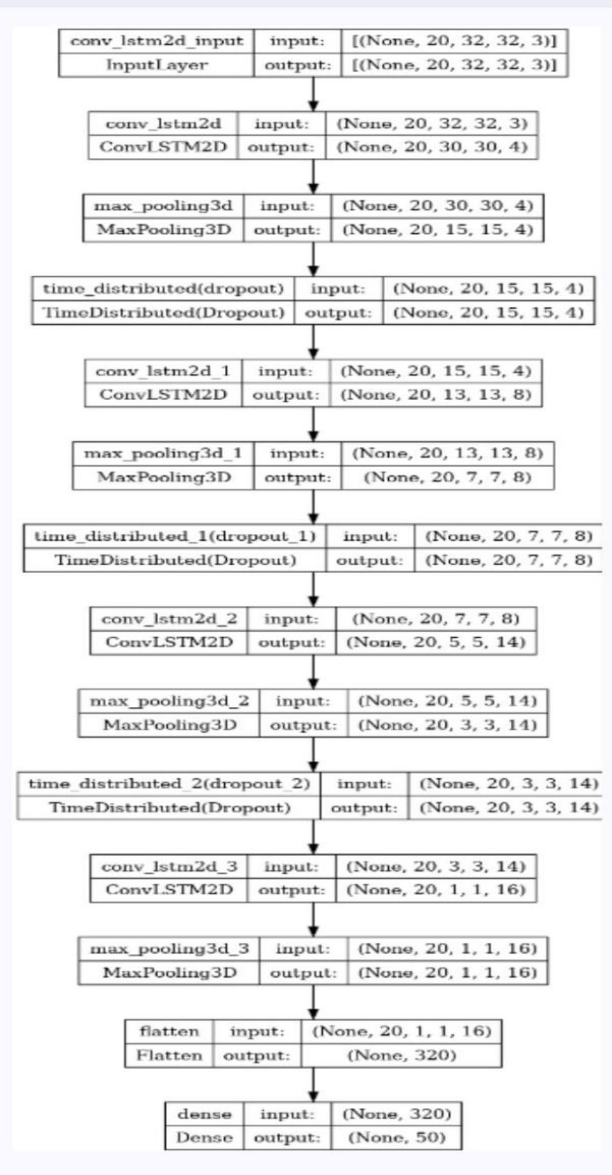


Fig.4 2D CNN LSTM Model Architecture

$$n_{out} = \left[\frac{n_{in} + 2p - k}{s}\right] + 1 \quad (2)$$

$$i_t = \sigma(x_t U^I + h_{t-1} W^i) \quad (3)$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f) \quad (4)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o) \quad (5)$$

$$\tilde{C}_t = \tanh\,(x_t U^g + h_{t-1} W^g) \quad (6)$$

$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t \quad (7)$$

$$h_t = \tanh(C_t) * o_t \quad (8)$$

- **Long-term Recurrent convolutional Network LRCN**

This model is built on 2DCNN and LSTM layers, each layer of which is time distributed. It also includes 4 convolutional layers, 4 max-pooling layers (2D time distributed layers), 3 dropout layers with a 25% dropout rate, and a dense layer serving as the output layer with a SoftMax output function [16]. Model compile with Adam's optimizer; metrics are accuracy, categorical cross entropy is the loss; epochs are 100; batch size is 4; and the validation split is 20%.

- **Creating a Transfer Learning Model**

When applying training over new data, evaluate the performance of the transfer learning model with new data videos to classify and predict the classes of a new dataset. After training the LRCN and 2 DCNN LSTM model, save weights and load them in new cells of code. Consider the same process of training with the same number of classes but the dataset class type was changed.

4. RESULT & DISCUSSION

A discussion of the outcomes of the deep learning models is presented in this section. In this work, two models are implemented, including 2DCNN LSTM & LRCN, & the metrics are Accuracy or Loss for Human activity Recognition.

1) Accuracy

A way to evaluate the effectiveness of a classification model is accuracy. Typically, a percentage is used to express it. Accuracy is defined as the percentage of forecasts when the anticipated value and actual value are equal. It is a true/false binary value for a particular sample. Graphing and tracking accuracy during the training process are common, even if the number is frequently connected to the overall or final model accuracy. Accuracy may be measured more easily than loss.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (9)$$

2) Loss

A loss function sometimes called a cost function, determines a forecast's likelihood or level of uncertainty depending on how far it differs from the true value. We can now see the model's performance in greater detail. Loss is the sum of all errors created for each sample in training and validation sets, as opposed to accuracy, which is expressed as a percentage.

$$Loss = -\frac{1}{m}\sum_{i=1}^{m} yi \cdot log(\widehat{yi}) \quad (10)$$

Table.1 Performance Evaluation of Models

Model	Training Acc	Training Loss	Validation Acc	Validation Loss
2DCNNLSTM	98.34	0.04	54.41	0.244
LRCN	99.08	0.02	66.91	0.162
Transfer 2DCNNLSTM	100	0.003	47.62	0.384
Transfer LRCN	100	0.001	69.39	0.198

The performance evaluation of the 2DCNNLSTM, Transfer 2DCNNLSTM, LRCN, and Transfer [27] LRCN is shown in Table.1, with the maximum Training Accuracy and Validation Accuracy being achieved by the Transfer LRCN at 100 and 69.39, respectively. Lowest Validation Loss of 0.162 and Lowest Training Loss of 0.001 respectively received via Transfer LRCN. The lowest 2DCNNLSTM validation accuracy is 47.62, whereas the lowest training accuracy is 98.34. Transfer LRCN successfully improved 2DCNN LSTM by 2%.

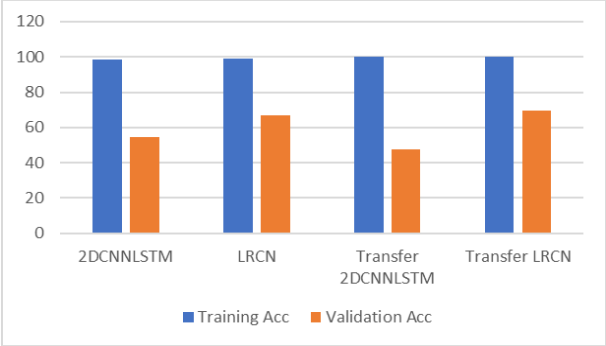


Figure 4: Accuracy Graph of Models

The maximum Training Accuracy and Validation Accuracy were attained by the Transfer LRCN at 100 and 69.39, respectively, in Fig.5, which depicts the accuracy graph of deep learning and transfer learning models such as 2DCNNLSTM, LRCN, Transfer 2DCNNLSTM, and Transfer LRCN.

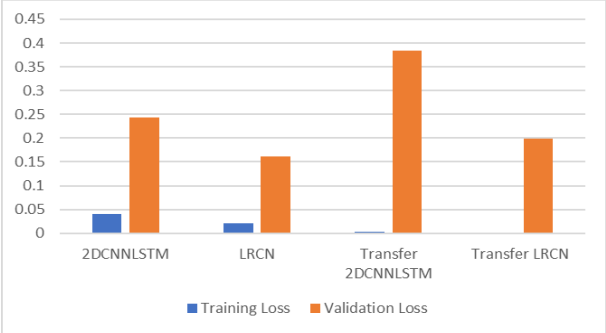
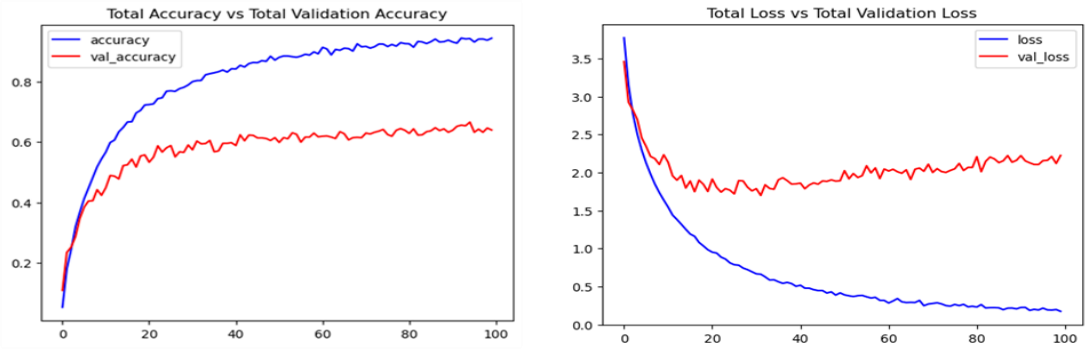


Figure 5: Loss Graph of Models

The loss graphs of deep learning and transfer learning algorithms such as 2DCNNLSTM, LRCN, Transfer 2DCNNLSTM, and Transfer LRCN are shown in Fig. 6. Lowest Validation Loss of 0.162 & Lowest Training Loss of 0.001 was obtained by Transfer LRCN, respectively. The lowest training accuracy of the 2DCNNLSTM is 98.34, while the lowest validation accuracy is 47.62.



(e) (f)

Figure 6: Accuracy(e) and Loss(f) of Transfer LRCN

Fig.7 is a graphical depiction of the Best Performed Model's LRCN Transfer Loss and Accuracy, with accuracy considered to be represented by the color blue and validation accuracy considered to be represented by the color red. In (f), loss is represented by the color blue, and validation loss is represented by the color red.

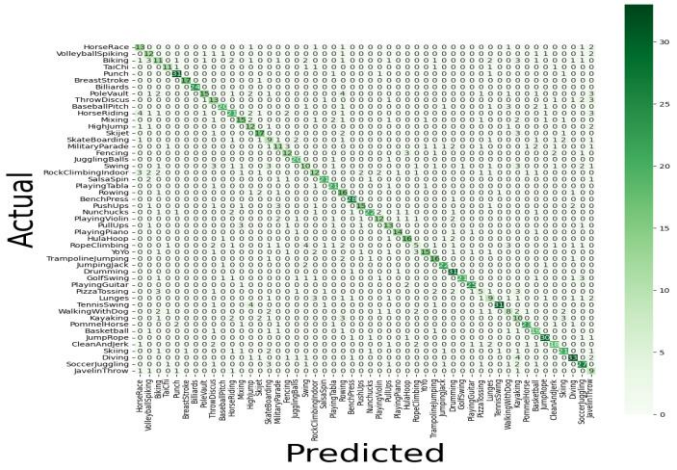


Figure 7: Confusion Matrix of Actual and Predicted variable

- **Research Gap**

Research on Human Activity Recognition (HAR) through video surveillance has become popular, particularly in inpatient rehabilitation and mobile health monitoring. There is a notable gap in effectively adapting Human Activity Recognition (HAR) classifiers to new users due to variations in activity patterns that can impact accuracy. The paper presents various deep learning models such as LRCN, Transfer LRCN, 2DCNNLSTM, and LRCN, emphasizing the superior performance of Transfer LRCN in terms of high accuracy scores. Further study is required to improve flexibility and efficiency in real-world situations, notwithstanding current improvements.

- **Contribution of the study**

This study makes an important contribution to the advancement of Human Activity Recognition (HAR) using video surveillance, especially in hospital environments. The research addresses the difficulty of adapting HAR classifiers to new users by introducing and evaluating deep learning models including LRCN, Transfer LRCN, and 2DCNNLSTM. The remarkable achievement of Transfer LRCN is shown in its exceptional performance, achieving the top scores for Training and Validation Accuracy. This discovery is essential for practical uses, guaranteeing reliable identification of activities even in the presence of various user behaviours. The study's findings will facilitate the improved utilization of HAR in inpatient rehabilitation, activity recognition, and mobile health monitoring, ensuring better accuracy and flexibility.

- **Limitation**

Although this study has made progress, it nevertheless has constraints. Relying on current data for offline training could impede the ability to adjust to distinct user behaviour patterns. The study mainly concentrates on a certain group of deep learning models, possibly overlooking other effective methods. Furthermore, the great accuracy of Transfer LRCN may not always be applicable in real-world situations characterized by dynamic and unpredictable behaviours. The study focuses on specific models and does not extensively investigate wider factors like computational costs and real-time processing capabilities. These constraints underscore the necessity for a thorough and pragmatic strategy to tackle the intricacies of various user behaviours in future study.

5. CONCLUSION

Human activity recognition (HAR) using video surveillance has recently witnessed an increase in popularity across the board, including inpatient rehabilitation, activity recognition, and mobile health monitoring. Activities recognise by videos, before being applied to new users, a HAR classifier is often trained offline with known users. The accuracy of this strategy may be compromised if the activity patterns of new users differ from those in the training data. It is impractical to create mobile applications from scratch because of the high cost of computing and the extensive learning curve for new users. The LRCN, Transfer LRCN, 2DCNNLSTM, and LRCN models were among the deep learning or transfer learning models proposed in this study. With scores of 100 and 69.39, the Transfer LRCN had the greatest Training Accuracy and Validation Accuracy. The Lowest Validation Loss and Lowest Training Loss, each of which was received via Transfer LRCN, were 0.16 and 0.001 respectively. While the 2DCNNLSTM's lowest validation accuracy is 47.62, its lowest training accuracy is 98.34. 2% of the 2DCNN LSTM was successfully increased by transfer LRCN.

Refrences

[1] N. Halim, "Stochastic recognition of human daily activities via hybrid descriptors and random forest using wearable sensors," *Array*, vol. 15, no. May, p. 100190, 2022, doi: 10.1016/j.array.2022.100190.

[2] Shruthi, P. Pattan, and S. Arjunagi, "A human behavior analysis model to track object behavior in surveillance videos," *Meas. Sensors*, vol. 24, no. August, p. 100454, 2022, doi: 10.1016/j.measen.2022.100454.

[3] J. Yang, Y. Xu, H. Cao, H. Zou, and L. Xie, "Deep learning and transfer learning for device-free human activity recognition: A survey," *J. Autom. Intell.*, vol. 1, no. 1, p. 100007, 2022, doi: 10.1016/j.jai.2022.100007.

[4] L. Zhu and L. Liu, "3D Human Motion Posture Tracking Method Using Multilabel Transfer Learning," *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/2211866.

[5] "Convolutional Neural Network (CNN) in Machine Learning - GeeksforGeeks." https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-learning/ (accessed Apr. 25, 2023).

[6] "An Overview of Deep Belief Network (DBN) in Deep Learning." https://www.analyticsvidhya.com/blog/2022/03/an-overview-of-deep-belief-network-dbn-in-deep-learning/ (accessed Apr. 25, 2023).

[7] A. Hussain, T. Hussain, W. Ullah, and S. W. Baik, "Vision Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance Videos," *Comput. Intell. Neurosci.*, vol. 2022, no. 1, 2022, doi: 10.1155/2022/3454167.

[8] D. Sun, J. Zhang, S. Zhang, X. Li, and H. Wang, "Human Health Activity Recognition Algorithm in Wireless Sensor Networks Based on Metric Learning," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/4204644.

[9] L. Qiao and Q. H. Shen, "Human Action Recognition Technology in Dance Video Image," *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/6144762.

[10] A. Mihoub, "A Deep Learning-Based Framework for Human Activity Recognition in Smart Homes," *Mob. Inf. Syst.*, vol. 2021, 2021, doi: 10.1155/2021/6961343.

[11] T. George Karimpanal and R. Bouffanais, "Self-organizing maps for storage and transfer of knowledge in reinforcement learning," *Adapt. Behav.*, vol. 27, no. 2, pp. 111–126, Apr. 2019, doi: 10.1177/1059712318818568.

[12] S. Li, J. Fan, P. Zheng, and L. Wang, "Transfer Learning-enabled Action Recognition for Human-robot Collaborative Assembly," *Procedia CIRP*, vol. 104, no. March, pp. 1795–1800, 2021, doi: 10.1016/j.procir.2021.11.303.

[13] W. Lao, J. Han, and P. H. N. De With, "Flexible human behavior analysis framework for video surveillance applications," *Int. J. Digit. Multimed. Broadcast.*, vol. 2010, 2010, doi: 10.1155/2010/920121.

[14] J. Sun, Y. Fu, S. Li, J. He, C. Xu, and L. Tan, "Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors," *J. Sensors*, vol. 2018, no. 1, 2018, doi: 10.1155/2018/8580959.

[15] Y. Y. Zhu, Y. Y. Zhu, Z. K. Wen, W. S. Chen, and Q. Huang, "Detection and recognition of abnormal running behavior in surveillance video," *Math. Probl. Eng.*, vol. 2012, 2012, doi: 10.1155/2012/296407.

[16] "Brief Review — LRCN: Long-term Recurrent Convolutional Networks for Visual Recognition and Description | by Sik-Ho Tsang | Medium." https://sh-tsang.medium.com/brief-review-lrcn-long-term-recurrent-convolutional-

networks-for-visual-recognition-and-9542bc7e8a79 (accessed Apr. 25, 2023).

[17] S. M. Hejazi and C. Abhayaratne, "Handcrafted localized phase features for human action recognition," *Image Vis. Comput.*, vol. 123, p. 104465, 2022, doi: 10.1016/j.imavis.2022.104465.

[18] X. Cui and R. Hu, "Application of intelligent edge computing technology for video surveillance in human movement recognition and Taekwondo training," *Alexandria Eng. J.*, vol. 61, no. 4, pp. 2899–2908, 2022, doi: 10.1016/j.aej.2021.08.020.

[19] R. Mar-Cupido, V. García, G. Rivera, and J. S. Sánchez, "Deep transfer learning for the recognition of types of face masks as a core measure to prevent the transmission of COVID-19," *Appl. Soft Comput.*, vol. 125, p. 109207, 2022, doi: 10.1016/j.asoc.2022.109207.

[20] V. Sarveshwaran, I. T. Joseph, M. Maravarman, and P. Karthikeyan, "Investigation on Human Activity Recognition using Deep Learning," *Procedia Comput. Sci.*, vol. 204, pp. 73–80, 2022, doi: 10.1016/j.procs.2022.08.009.

[21] P. Kumar and S. Suresh, "FLAAP: An Open Human Activity Recognition (HAR) Dataset for Learning and Finding the Associated Activity Patterns," *Procedia Comput. Sci.*, vol. 212, no. C, pp. 64–73, 2022, doi: 10.1016/j.procs.2022.10.208.

[22] "In machine learning, when is one hot encoding better than target (mean) encoding? Why would you ever use OHE over target encoding? - Quora." https://www.quora.com/In-machine-learning-when-is-one-hot-encoding-better-than-target-mean-encoding-Why-would-you-ever-use-OHE-over-target-encoding (accessed Apr. 25, 2023).

[23] "sklearn.preprocessing.OneHotEncoder — scikit-learn 1.2.2 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html (accessed Apr. 25, 2023).

[24] "CRCV | Center for Research in Computer Vision at the University of Central Florida." https://www.crcv.ucf.edu/data/UCF50.php (accessed Apr. 25, 2023).

[25] "5. CNN-LSTM — PseudoLab Tutorial Book." https://pseudo-lab.github.io/Tutorial-Book-en/chapters/en/time-series/Ch5-CNN-LSTM.html (accessed Apr. 25, 2023).

[26] S. Anoopa, A. Salim, and S. Nadeera Beevi, "Advanced video anomaly detection using 2D CNN and stacked LSTM with deep active learning-based model," *Kuwait J. Sci.*, vol. 49, Jun. 2022, doi: 10.48129/KJS.SPLML.19159.

[27] "Long-term Recurrent Convolutional Network for Video Regression | by Alexander Golubev | Towards Data Science." https://towardsdatascience.com/long-term-recurrent-convolutional-network-for-video-regression-12138f8b4713 (accessed Apr. 25, 2023).