International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307 (Online), Volume-14 Issue-2, May 2024

Human Action Recognition using Long Short-Term Memory and Convolutional Neural Network Model

Shreyas Pagare, Rakesh Kumar

Abstract: Human Action Recognition (HAR) is the difficulty of quickly identifying strenuous exercise performed by people. It is feasible to sample measures of a body's tangential acceleration and speed using inertial sensors and train them to learn model skills of correctly categorising behaviour into relevant categories. In detecting human activities, the use of detectors in personal and portable devices has increased to understand better and anticipate human behaviour. Many specialists are working toward developing a classification that can distinguish between a user's behaviour and raw data while utilising as few resources as possible. A Long-term Recurrent Convolutional Network (LRCN) is proposed as a comprehensive human action recognition system based on deep neural networks in this paper.

Keywords: Human Action Recognition, Convolutional Neural Network, Long Short-Term Memory, Long Short-Term Memory

I. INTRODUCTION

Human action recognition (HAR) is an imperative aspect of today's generation since it can analyse novel information about human actions from raw data. As a result of the rise of interpersonal communication applications, HAR innovation has emerged as a key study field both locally and internationally. Individuals can categorise any type of urban transportation and gather the knowledge that the body requires to function effectively by conveying information from ordinary items, setting the framework for future applications [1].

Observing human behavior is essential in interpersonal interactions and intimate communication. It's tough to retrieve because this includes details about specific individuals, their attitudes, and mental factors. In the research domain of computational intelligence, one of the key objects of analysis is indeed the individual's ability to perceive another person's behaviour. A fundamental element of several consumer items was action recognition [2]. For example, game systems like the Nintendo Wii and Microsoft Kinect utilise gesture recognition or full-body motion tracking to enhance the gaming experience significantly. While created for the entertainment sector, these systems have found new applications in fields such as personal fitness training and rehabilitation, as well as driving new research in action recognition.

Manuscript received on 13 July 2023 | Revised Manuscript received on 10 May 2024 | Manuscript Accepted on 15 May 2024 | Manuscript published on 30 June 2024.

**Correspondence Author(s)*

Shreyas Pagare*, Research Scholar, Department of Computer Science & Engineering, RNTU University, Bhopal (M.P.), India. E-mail: shreyas_au211443@aisectuniversity.ac.in, ORCID ID: 0009-0001-8147-2389

Dr. Rakesh Kumar, Research Guide, Department of Computer Science & Engineering, RNTU University, Bhopal (M.P.), India. E-mail: rakeshmittan@gmail.com

Finally, some sporting goods, such as the Philips Direct Life or Nike+ running shoes, contain motion sensors and provide feedback to both amateur and professional athletes.

All of these instances demonstrate the importance of recognizing human activities in both academia and industry. Developing HAR systems that match application and user needs remains a problematic issue, despite significant progress in inferring activities from on-body encoders and prototyping and deploying action recognition systems. This is true even if HAR approaches that worked well for one recognition challenge are applied to a different problem domain.

A. Machine Learning

A Machine Learning algorithm is a sequence of instructions and empirical methods for extracting relevant insights from respondents and internal restructuring based on them. It is the overarching theme of a Machine Learning model. A model is the most essential subset of machine learning. A Machine Learning Algorithm is employed as a model. To achieve the desired result, an approach integrates all the assumptions that a model is designed to deliver, relying on the provided information. A Machine Learning procedure occurs when massive quantities of data are supplied to the device and then used to train an algorithm to discover hidden actionable insights, leveraging this fact. These observations are then utilized to make a Training Data that addresses the problems using a method [3].

B. Deep Learning

Deep Learning is a machine learning subfield that understands the world as a recursive hierarchical order of principles, with each separate dimension connected to smaller units and many more conceptual representations deduced from aspects of less tangible ones. It achieved excellent scalability and performance by learning to reflect reality as a recursive hierarchical system of notions. Deep learning networks learn by detecting complicated patterns from the data they receive. The systems can establish numerous areas of granularity to make sense of the data by establishing simulation tools that are composed of many convolutional filters [4].

A deep neural network, a category of deep learning models, can be trained using a substantial array (in the millions) of images, including those depicting living beings. Essentially, this form of neuron adapts based on the cells it obtains from the visuals. It can distinguish and label collections of data that resemble distinct qualities, including hands, heads, and ears, which confirm the existence of a human in a vision.

II. LITERATURE REVIEW

The method of identifying body gestures or motions and then predicting or defining stages of action or behavior is known as human action recognition (HAR) [3]. Its many applications, including the Army care system, athletic rehabilitation following a disability or injury, and medical trials for abnormality therapy, stimulate industrial and academic interest in future research and development. With the pervasive use of handheld personal digital devices like smartphones, smartwatches, and multi-media stations that create a wide range of different forms of chronic data, such as recording devices, photograph streams, and geographic timbers, there is going to be a considerable Personalization based on human action recognition, especially in national healthcare [2] which elucidate a deep learning-based HAR strategy Researchers use an accelerometer pulse to build a power spectrum visual, which they then inject into a convent. The feature extraction step is effectively replaced by the spectrogram synthesis phase, which adds some initial cost to the network training [3]. Taking a pure accelerometer sensor as the input to a convolutional neural network (CNN), perform a 1-D inversion on every sample. The spatial linkages between distinct sensor components may be lost as a result of this method. They concentrate on datasets that are freely available and originate mainly from embedded sensors (such as phones) or monitoring devices [5]. adopt a comparable tactic, Researchers have used the same existing data, and have used two-dimensional inversion to express dynamic impulses in a specific cable during their investigation [6] This uses an inter multilayer neural which incorporates linear accelerating and rotational mobility signals to categorize usual tasks using a labeled database comprising topmost motions for implementation of CNNs for the occupation classification task. Because the sorting assignment participants undertake is highly individual, data indicators collected from each respondent are being used to train various learning models [8] to build a monitoring system out of six inertial measurement devices. Following network analysis, to create a functional model, the researchers identified several performance parameters that survived the statistical method and then employed the random forest (RF) classifier to categorise the events. Finally, the correctness level was 84.6 %. The study described a haptic feedback inertial sensors action recognition system and its submission in healthcare diagnosis [6]. For feature selection, Reprieve and consecutive advance drifting scan (RCADS) were coupled. Finally, for action classification and comparison, the approaches of Naive Bayesian and k-nearest neighbour (KNN) have been employed. ML algorithms can heavily rely on heuristics for extracting subjective features, particularly in tasks related to detecting people's daily movements. The knowledge base possessed by humans is usually a major hurdle [9]. Deep learning techniques, which can automatically generate key features from telemetry data even during the preprocessing step, and substantially reduce the original descriptors while elevating conceptual patterns, have been used to resolve this difficulty. Adapting deep neural networks to the scope of human wearable sensors is indeed a recent research issue in analytical thinking [19], considering their effectiveness in classification, speech synthesis, computational linguistics, as well as other areas. Zens et al. have suggested translating three-axis sensor readings into a "picture" representation, then employing a CNN with three convolutions and one fully connected network to represent human movements. Consultation of additional literature available for the researcher [18] has been done to improve the perception of the broad field.

III. RESEARCH METHODOLOGY

Deep Learning is used in this research to construct a system that recognizes human action. For the implementation of human action recognition on videos, the proposed system utilises a convolutional neural network paired with an extended short-term memory Network. In TensorFlow, we'll need to use two distinct structures and approaches. Finally, we'll adapt the successful system to YouTube video estimations [4].

A. Image Classification

Image classification is where a computer can examine an image to see which 'class' it belongs to. (Or there's a chance it's a 'class.') A class is simply a phrase, such as 'vehicle,' 'animal, ' and so on. Pass an image into a filter (whether it's a learned deep neural network (convolutional neural network or multi-layer perceptron) or a conventional classifier) to get a class prediction. Let's say you want to insert a human photo. The computer analyses the image and determines whether it is of a male or female. (Or the possibility that it's a female.) [15][16]

A video is simply a series of static images (known as frames) that are updated very quickly to provide the impression of motion. Take a look at the footage of a cat jumping on a bookcase below. (Converted to .gif format), which is simply a compilation of 15 separate still images that are updated one after the other [19].

B. Video Classification

Video classification is the process of applying a tag to a central aspect of its pixels. A competent YouTube clip predictor does far more than just provide precise panel tags; so, it reflects the real clip that uses the elements and tags of the consecutive frames [7]. For example, a clip might show a shrub inside one shot; however, the major championship (e.g., "hiking") could be something wholly distinct. The objective establishes the level of precision required for the identifiers used to characterise the pieces and clips. Typical tasks include providing one or even more worldwide tags to a clip, and also one or maybe more descriptions for each panel from the inside of the clip [10].

C. Convolutional Neural Network

A convolutional neural network (CNN or ConvNet) is a multi-layer perceptron designed to operate on visual data and specialises in vision estimation and forecasting. It operates with seeds (called filters) that are placed over the portrait and simply create a saliency map (which indicates how well a particular component is visible at a precise point in the image or not). Thus, it initially creates a comparatively small number of parameters.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

However, as we proceed with greater depth into the intranet, this same number of nodes develops as well as the magnitude of graphs gets smaller without having to lose pertinent data utilizing accumulating processes [18].

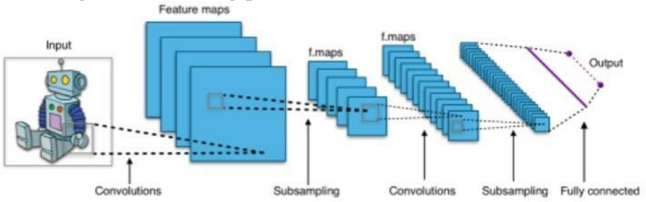


Figure. 1: Convolutional Neural Network

The layers of a ConvNet learn features of increasing complexity, such as detecting shapes in the first layer and recognising diverse postures in the last layer.

D. Long Short-Term Memory (LSTM)

Because it considers all various sources in order when generating output, a long short-term memory system was developed specifically to accommodate a data series. Although LSTMs are a subtype of RNN (Recurrent Neural Network), RNN has been demonstrated to be inadequate in dealing with long-standing dependencies in input sequences, induced by the Disappearing Gradient problem. To overcome the disappearing slope and allow an LSTM unit to recall the context of long input data, LSTMs were developed [11].

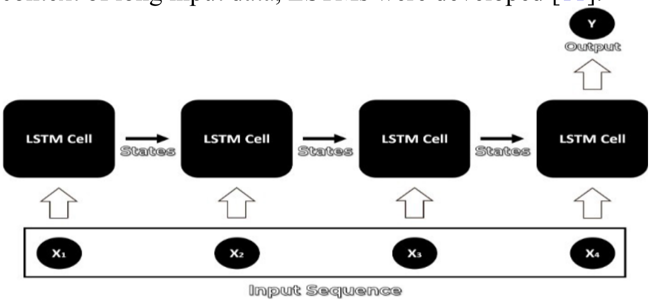


Figure 2: Long Short-Term Memory

This enhances an LSTM's ability to handle sequential data problems, including time series forecasting, speech synthesis, dictionary lookup, and orchestral composition. But for now, let's focus on how LSTMs can help us build more powerful action recognition models. We'll now examine how Action Recogniser will be implemented. We will utilize a deep convolution network (DCN) + long short-term memory (LSTM) Network to undertake Action Recognition when leveraging the geographic element of the clips [19].

E. Recurrent Neural Network (RNN)

Recurrent neural networks identify consecutive features of the information and anticipate the following likely situation using patterns. RNNs have been utilised in deep learning and the development of algorithms that mimic neuronal behaviour in the nervous system. They're incredibly beneficial in cases where comprehension is essential to predicting a response. They differ from other neural networks in that they use responses to examine a set of memories that influence accurate results. Such natural cycles and good documentation are needed to endure. The recall is a common term for this phenomenon [12].

Prominent instances of recurrent neural network usage include word embeddings, where the subsequent word in a term and the next word in a sentence are dependent on the facts that came before. Another innovative experiment was using a Recurrent neural network that received training on Literary plays and created Shakespeare-like prose. Computational inventiveness is a sort of RNN-based literature. This reproduction of artistic expression is facilitated by the AI's linguistic knowledge and interpretation, which are obtained during its training phase.

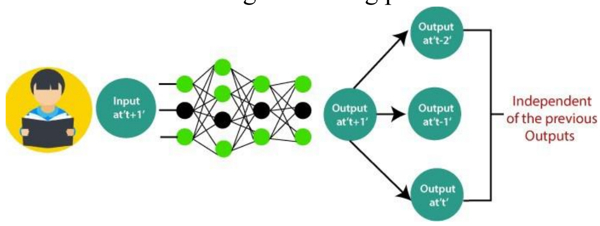


Figure 3: Recurrent Neural Networks [18]

F. LSTM vs RNN

The extended short-term memory (LSTM) network is a type of recurrent neural network that includes both special and conventional subunits. There is, in fact, a 'memory block' in LSTM units that can store information for extended periods. This memory cell enables them to learn long-term dependence. The main distinction between RNN and LSTM is that LSTM retains information in memory for a lengthy period. In this scenario, LSTM outperforms RNN because it can keep information in memory for a more extended period [13].

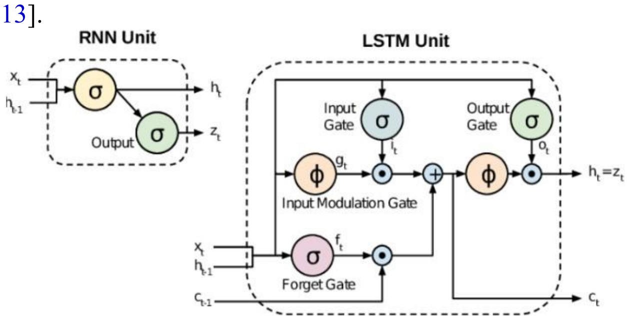


Figure 4: LSTM vs RNN [16]

IV. PROPOSED MODEL

A. CNN + LSTM

A CNN would be used to retrieve distance measures at a specific sampling interval inside the original signal (clip), and an LSTM would be used to discover and predict future performance among panels.

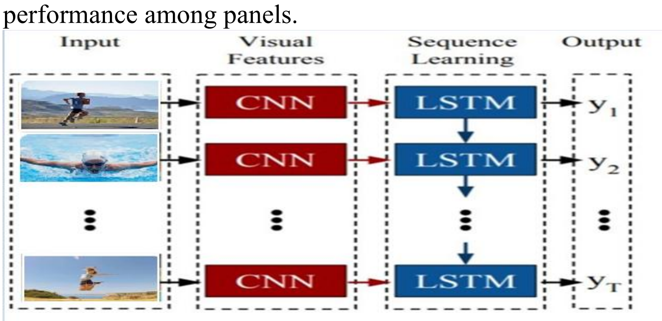


Figure 4: CNN and LSTM [14]

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

These two structures we are using to combine CNN and LSTM are as follows:
1 ConvLSTM
2 LRCN
TensorFlow can be used for each of these approaches.

B. Conv LSTM

ConvLSTM is a recurrent neural network that features multilayer properties within both insight and state-to-state stages, which anticipate spatial patterns. The ConvLSTM forecasting condition of the unit within the grid utilises the supplies and previous conditions of its local neighbours. A simple way to achieve this is by using a transformation function in the state-to-state and insight conversions. The following are indeed the ConvLSTM fundamental equations, where '*' represents the convolution operator and 'o indicates the Hadamard product:

$$it = \sigma (Wxi * Xt + Whi * Ht-1 + Wci \circ Ct-1 + bi)$$
$$ft = \sigma (Wxf * Xt + Whf * Ht-1 + Wcf \circ Ct-1 + bf)$$
$$Ct = ft \circ Ct-1 + it \circ \tanh (Wxc * Xt + Whc * Ht-1 + bc)$$
$$ot = \sigma (Wxo * Xt + Who * Ht-1 + Wco \circ Ct + bo)$$
$$Ht = ot \circ \tanh (Ct)$$

We should anticipate a ConvLSTM with a larger transition unit to detect greater oscillations and a smaller unit to identify delayed actions, if we interpret its transitions as disguised descriptions of moving objects.

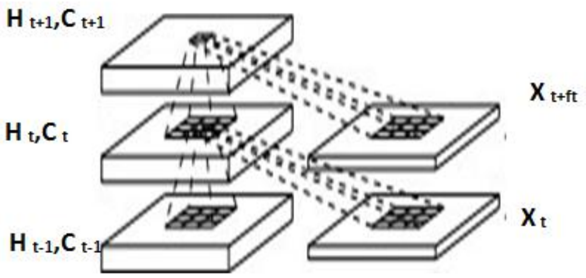


Figure 6: Inner Structure of Conv LSTM

Before executing the convolutional operation, padding is essential to ensure that the results have the same number of data points as the input. Determining the aspect of the outside globe could be equivalent to padding the disguised conditions at the border crossing points. Before the first entry, we generally set all the variables of the LSTM to zeroes, which essentially translates to "profound ignorance" of the long term.

C. LRCN

A model for activities that involve sequence information (inputs or outputs), perhaps optical, verbal, or other, that blends a fully convolutional graphical extracted features (such as a CNN) with such a system that really can be taught to identify and manufacture time evolution. The core of our approach is depicted in Figure 7.

The addition of new features φV (.) with variables V, almost always a CNN, is applied to every realizing xt (a single picture or video frame) to produce a repaired generative model φV. (xt). The results of V are therefore already in modules that learn recurring sequences.

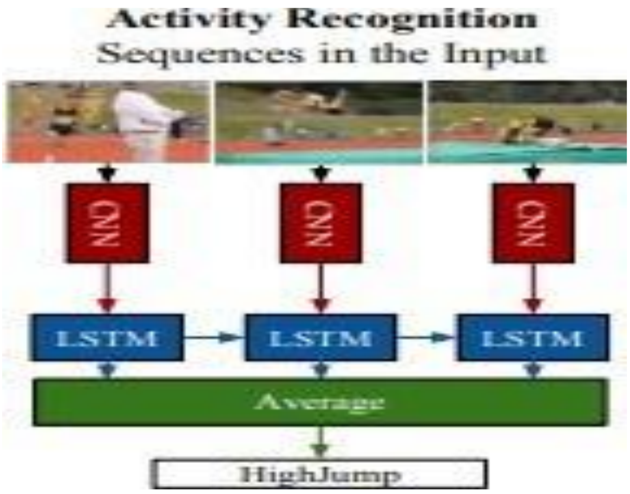


Figure 7: LRCN Model for Action Recognition

D. Data Collection

UCF50 is a data set of 50 action categories for action recognition culled from real-life action videos on YouTube. This set of data serves as a supplement to the Vimeo Action set of data (UCF11), which contains 11 different action categories. The commonality of the action recognition numbers sets available is unrealistic and staged by actors. Our primary objective in gathering data is to provide the computer image community with a source of action recognition data drawn from realistic YouTube videos. Our dataset is highly demanding due to significant changes in camera activity, object appearance and posture, object ruler, perspective, disordered backdrop, lighting conditions, and other parameters.

V. MODEL CREATION

A. Conv LSTM Approach

We combine ConvLSTM cells to execute the first approach. Inside the system, the ConvLSTM unit is a variant of the LSTM unit that incorporates convolutional procedures. It was an LSTM with a convolution built in, enabling it to distinguish spatial properties as information, even while considering future performance predictions. Regarding video identification, this method numerically captures the relationship between individual frames in terms of both space and time. While a regular LSTM can only handle one dimension of input, due to its convolution structure, the ConvLSTM can handle three dimensions (width, height, and number of channels). As such, it is not appropriate for modelling spatiotemporal data on its own [17].

B. Model of ConvLSTM

Recurrent layers from Keras ConvLSTM2D will be used to construct the model. Additionally, the filter and kernel sizes required for multilayer processes are considered by the ConvLSTM2D layer. Finally, a deep network with a linear kernel is fed the flattened layer output, which provides a frequency for every action category. We'll also utilise dropout layers to prevent overfitting of the model on the data and a MaxPooling3D layer to reduce the size of the frames and avoid unnecessary calculations.

Retrieval Number: 100.1/ijsce.I96970812923
DOI: 10.35940/ijsce.I9697.14020524
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307 (Online), Volume-14 Issue-2, May 2024

The design is straightforward, with only a few trainable parameters. This is because we are only dealing with a small portion of the information, which does not need the adoption of a large-scale model [18].

C. LRCN Approach

To apply the LRCN Approach, we'll merge the Pooling layer and LSTM layers into a single component in this stage. Similarly, a Convolution layer and a hidden layer prototype can be used individually. A pre-trained system, ideally suited for this purpose, could be used to retrieve spatial features from video sequences to utilise the Network model. The LSTM model could then use the Feature representation to forecast the action taken inside the clip.

However, the Long-term Recurrent Convolutional Network (LRCN), which integrates CNN and LSTM layers into a single model, will be used in this case. The LSTM layer(s) are tasked with temporal sequence modelling after the Convolutional layers have extracted spatial properties from the frames at each time step. The network can acquire spatiotemporal features through end-to-end training, producing a reliable model.

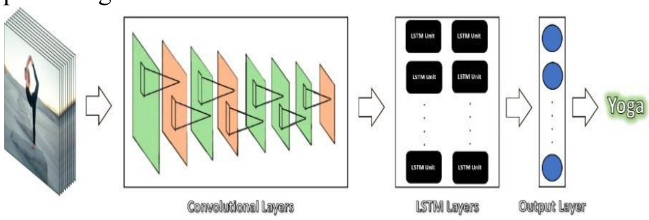


Figure 9: LRCN Approach

A Time-Distributed container component would also be used, enabling us to deploy the same surface to each pixel of the clip independently. As just a corollary, if indeed the gradient intake structure were (breadth, elevation, number of streams), it can additionally receive information of pattern (number of images, spacing, size, number of channels), which is tremendously helpful as it enables users to enter the entire movie into the method in one go [19].

D. LRCN Model

We will utilise time-circulated Conv2D layers, MaxPooling2D, and Dropout layers to develop the LRCN architecture. The feature gathered from the Conv2D layers will be flattened and sent to an LSTM layer using a flattening surface. The output from the LSTM layer would then be used by the Thick layer, which uses hidden neurons, to forecast the operation that would be carried out.

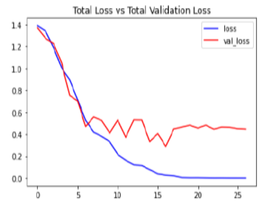
E. LRCN Model Evaluation

We'll analyse the model using testing data after it has been trained, and the accuracy rate is 92%, with a loss rate of 22%.

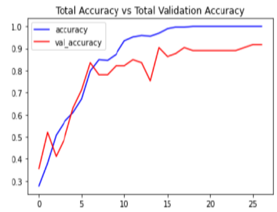
VI. RESULTS

In this paper, we have employed two different approaches: a grouping of convolutional neural networks and a long short-term memory (LSTM) model. The First approach we used is ConvLSTM, and the second approach is LRCN. Here is the visualization of the ConvLSTM

Loss of ConvLSTM model

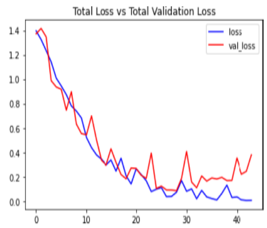


Accuracy of ConvLSTM model

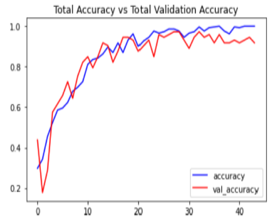


Here is the visualization of the LRCN Approach

Loss of LRCN model



Accuracy of LRCN model



Approach	Loss	Accuracy
ConvLSTM	89%	80%
LRCN	22%	92%

The Conv LSTM model achieves an accuracy rate of 80%, while the LRCN model achieves an accuracy rate of 92%. As we can see, the LRCN gives better accuracy compared to the Conv LSTM model. Therefore, we will utilise the LRCN Model to perform predictions on YouTube videos.

VII. CONCLUSION

Human action recognition has a variety of applications due to its impact on people's pleasure. It has become a significant tool in personalised medicine, including the prevention of weight gain and elderly care. In this paper, we have used a convolutional neural network with long short-term memory. We tried two different approaches. First, we create a model using ConvLSTM, and a second model, using LRCN and TensorFlow, is used for each of these approaches. We have used the UCF50 dataset for model creation. UCF50 is a data set of 50 action categories for action recognition culled from real-life action videos on YouTube. The ConvLSTM approach achieves 80% accuracy with a 89% loss, while the LRCN approach yields 92% accuracy with a 22% loss. After examining the results of both approaches, we found that the LRCN approach yielded a better outcome compared to ConvLSTM. Therefore, the LRCN Model was used to perform predictions on YouTube videos.

DECLARATION STATEMENT

Funding	No, I did not receive it.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval or consent to participate, as it presents evidence that is not subject to interpretation.
Availability of Data and Materials	Not relevant.
Authors Contributions	I am the sole author of the article.

Retrieval Number: 100.1/ijsce.I96970812923
DOI: 10.35940/ijsce.I9697.14020524
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

REFERENCES

1. M. G. Morshed, T. Sultana, A. Alam, and Y. K. Lee, "Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities," Sensors, vol. 23, no. 4. MDPI, Feb. 01, 2023. doi: 10.3390/s23042182. https://doi.org/10.3390/s23042182
2. T. O. Araoye, E. C. Ashigwuike, A. C. Adeyemi, S. V. Egoigwe, N. G. Ajah, and E. Eronu, "Reduction and control of harmonic on three-phase squirrel cage induction motors with voltage source inverter (VSI) using ANN-grasshopper optimization shunt active filters (ANN-GOSAF)," Sci Afr, vol. 21, Sep. 2023, doi: 10.1016/j.sciaf. 2023 01785. https://doi.org/10.1016/j.sciaf.2023.e01785
3. K. D. Addo, F. Davis, Y. A. K. Fiagbe, and A. Andrews, "Machine learning for predictive modelling of the performance of an automobile engine operating without a coolant thermostat," Sci Afr, vol. 21, Sep. 2023, doi: 10.1016/j.sciaf.2023.e 01802. https://doi.org/10.1016/j.sciaf.2023.e01802
4. P. Lara-Benítez, M. Carranza-García, and J. C. Riquelme, "An Experimental Review on Deep Learning Architectures for Time Series Forecasting," Int J Neural Syst, vol. 31, no. 3, Mar. 2021, doi: 10.1142/S0129065721300011. https://doi.org/10.1142/S0129065721300011
5. E. C. Nwosu, G. N. Nwaji, C. Ononogbo, I. Ofong, N. V. Ogueke, and E. E. Anyanwu, "Effects of water thickness and glazing slope on the performance of a double-effect solar still," Sci Afr, vol. 21, Sep. 2023, doi: 10.1016/j.sciaf.2023.e 01777. https://doi.org/10.1016/j.sciaf.2023.e01777
6. S. Chae, J. Shin, S. Kwon, S. Lee, S. Kang, and D. Lee, "PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network," Sci Rep, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-91253-9. https://doi.org/10.1038/s41598-021-91253-9
7. H. W. Hounkpatin, H. E. V. Donnou, K. V. Chegnimonhan, G. H. Hounguè, and B. B. Kounouhewa, "Thermal characterisation of insulation panels based on vegetable Typha domengensis and starch," Sci Afr, vol. 21, Sep. 2023, doi: 10.1016/j.sciaf. 2023. e01786. https://doi.org/10.1016/j.sciaf.2023.e01786
8. S. Ghimire, Z. M. Yaseen, A. A. Farooque, R. C. Deo, J. Zhang, and X. Tao, "Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks," Sci Rep, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-96751-4. https://doi.org/10.1038/s41598-021-96751-4
9. H. A. Mupambwa et al., "A 2-year study on the spatio-temporal changes in trace metal concentrations in sediment, water and plants within the Walvis Bay Lagoon, Namibia," Sci Afr, vol. 21, Sep. 2023, doi: 10.1016/j.sciaf.2023.e01787. https://doi.org/10.1016/j.sciaf.2023.e01787
10. J. Donahue et al., "Long-term Recurrent Convolutional Networks for Visual Recognition and Description," Nov. 2014, [Online]. Available: http://arxiv.org/abs/1411.4389 https://doi.org/10.21236/ADA623249
11. O. G. Famutimi, I. O. Adewale, and K. R. Adegoke, "Inhibition characteristics of peptide extracts of four medicinal plants on activities of bovine trypsin," Sci Afr, vol. 21, Sep. 2023, doi: 10.1016/j.sciaf.2023.e01795. https://doi.org/10.1016/j.sciaf.2023.e01795
12. M. Shreyas Pagare and D. Rakesh Kumar, "Survey on Human Action Recognition System, Challenges and Applications," 2023.
13. M. O. Mario, "Human activity recognition based on single sensor square HV acceleration images and convolutional neural networks," IEEE Sens J, vol. 19, no. 4, pp. 1487–1498, Feb. 2019, doi: 10.1109/JSEN.2018.2882943. https://doi.org/10.1109/JSEN.2018.2882943
14. G. Theses and J. Pang, "Scholar Commons Human Activity Recognition Based on Transfer Learning," 2018. [Online]. Available: https://scholarcommons.usf.edu/etd
15. T. N. Sainath, O. Vinyals, A. Senior, and H. H. Sak, "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks."
16. S. Gopali, F. Abri, S. Siami-Namini, and A. S. Namin, "A Comparative Study of Detecting Anomalies in Time Series Data Using LSTM and TCN Models," Dec. 2021, [Online]. Available: http://arxiv.org/abs/2112.09293
17. Y. Kim, Y. Jernite, D. Sontag, and A. M. Rush, "Character-Aware Neural Language Models," Aug. 2015, [Online]. Available: http://arxiv.org/abs/1508.06615
18. O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and Tell: A Neural Image Caption Generator," Nov. 2014, [Online]. Available: http://arxiv.org/abs/1411.4555
19. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.04214

AUTHORS PROFILE

Shreyas Pagare is a Ph.D. scholar in the Department of Computer Science Engineering from Rabindranath Tagore University, Bhopal (M.P.). He completed his Master of Engineering (Computer Science & Engineering) from IET DAVV, Indore in 2012. His interests include Database Technologies, Data Structures, Data Mining, Data Warehousing, Artificial Intelligence, and Machine Learning. He has approximately 15 years of experience in teaching and Research at the college and University level. He has organised numerous trainings and Workshops. He has delivered multiple expert lectures in various organisations. He is working on various funded projects. He has published numerous research papers in Reputable Journals (i.e., SCOPUS and UGC-approved) and presented papers at various National and International Conferences. He is the author of Book Chapters and Patents.

Dr. Rakesh Kumar is an Associate Professor in the Department of Computer Science & Engineering and Head of the Centre for IoT & Advanced Computing. He is a member of the Institute Innovation Council (IIC) of the MHRD and the Innovation Ambassador at the University. He completed his Ph.D. in Computer Science & Engineering. His area of interest includes IoT, IoE, Software Engineering, Operating Systems, Database Technologies, Data Structures, Data Mining, Data Warehousing, Robotics, Drone Technology, Artificial Intelligence, and Computer Architecture. He has approximately 17 years of experience in teaching and Research at the college and University level. He has organised numerous trainings and Workshops. He has delivered multiple expert lectures in various organisations. He has organized AICTE funded FDP. He has guided numerous Projects at the State and National Level competitions. He is working on various funded

projects. He has published multiple research papers in Reputable Journals (i.e., SCI, SCOPUS, and UGC-approved) and presented many papers at various National and International Conferences. He is the author of Book Chapters and Patents. He has been appointed as an Editorial board member and reviewer in different journals and international conferences. He was recognize by "Quarterly Franklin Membership" (Membership ID#VY34049), London Journal of Engineering Research (LJER), London Journals Press (UK) for research paper "Inside Agile Family Software Development Methodologies. He is a founding member of the Centre for IoT & Advanced Computing.