Golden Sun-Rise
International Journal of Multidisciplinary on Science and Management
ISSN: 3048-5037/ Volume 1 Issue 4 December 2024 / Page No: 31-41
Paper Id: IJMSM-V1I4P106 / Doi:10.71141/30485037/V1I4P106

Original Article

Human Activity Recognition (HAR) System Using Deep Learning

Stephen Mkegh Nengem[1], Friday Haruna[2], Samuel Ayua[3]

[1,2]Department Mathematics and Statistics, Taraba State University, Jalingo, Nigeria.

[3]Department of Computer Science, Taraba State University, Jalingo, Nigeria.

Received: 12 October 2024 Revised: 23 October 2024 Accepted: 02 November 2024 Published: 18 November 2024

Abstract - *This research explores the application of deep learning in human activity recognition (HAR). As technological advancements continue, HAR plays a pivotal role in fields like healthcare, sports, and security. Leveraging deep learning models, particularly neural networks (NN) and convolutional neural networks (CNN), enhances the accuracy and efficiency of HAR systems. Hence, the article discretized the outline of the human constitution into various limited variables. At that point entire solidness network of the outline was determined with the guide of the use of a developed and trained human activity recognition model; Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and combined CNN-LSTM models. The frames were compared and predict the type of activity, classification and captioning were done in real time. The analysis of their performances were compared to get the optimal result.*

Keywords - *Human Activity Recognition (HAR), Convolution Neural Networks (CNN) Deep learning, Human Computer Interaction.*

I. INTRODUCTION

The increasing availability of affordable human motion tracking devices, such as the inertial measurement unit (IMU) or the depth camera, has propelled an emerging area of computer vision and ubiquitous computing that focuses on human activity recognition (HAR) from the sensor's data. The ability to recognize and understand human activities from a series of observations is important for applications such as healthcare and rehabilitation, entertainment, surveillance, and human-computer interaction. The term deep learning refers to neural networks with multiple hidden layers that can learn increasingly abstract representations of the input data. In contrast to traditional machine learning, one of the main advantages of deep learning is that, it is able to automatically learn features directly from the raw sensor data, avoiding the need for the manual extraction of features [1].

However, according to [2], a great deal of domain knowledge and intuition is often required to build a successful deep learning model for complex real-world problems such as HAR, for example in architectural choices (e.g., the number of hidden layers), data augmentation, or parameter tuning. Although deep learning has achieved very promising results in image and speech recognition, currently it is still an open research question regarding how effective deep learning techniques are for HAR as well as what potentially novel applications and insights can be brought within the HAR domain thanks to deep learning methods.

Human action recognition forms one of the most important application domains in computer vision. The main assignment here is to absolutely and effectively described human actions from a previously unseen data sequence acquired by sensors. The ability to recognize, understand and interprets complex human actions

can enable one construct many important applications such as intelligent surveillance systems, human-computer interfaces, health care and most importantly it will be useful to security and military applications [1].

Most current methods build classifiers based on complex handcrafted features computed from the inputs, which are driven by tasks and uncertain. [3] proposed a deep model convolutional neural network (CNN) in Matlab for human action recognition that can act directly on the raw inputs. In addition, an efficient pre-training strategy was introduced to reduce the high computational cost of kernel training to enable improved real-world applications. There is a rapid increase in recent years to the number of researchers and techniques focusing on human action recognition, and this has significantly improved its accuracy.

However, according to [2] action recognition is still a challenging problem due to many issues ranging from large intra-class difference, fuzzy boundary between classes, viewpoint variations, occlusion, appearance, computational time, influence of environments and recording settings, in particular from realistic videos. Moreover, to have a complete human action recognition system, we need a combination of several disciplines including psychology and ontology.

In an era of the smart city video monitoring is important to improve the quality of life and to ensure secure zones. For optimum area coverage, the places surveilled may have surveillance cameras mounted at a specific distance. This makes security system more reliable as well as it is necessary for best analysis and deep understanding of films. The video data driven system is beneficial for the healthcare, transportation, manufacturing, educational and retail sectors. Identifying the specific incident that the feed is to be dealing with, for example, suspicious activity at an airport, bus stop or train station is the goal of every camera feed [4]. When acknowledging human action is particularly desirable, these are the select few illustrative areas. With HAR based systems an alert of abnormal activity is raised to the control room. It is crucial to be aware of a few specific items in such cases instead of sitting in front of the camera feed and observing what is happening every second. A central objective of human activity recognition [5] is to predict actions and their interactions from an unheard of data sequence.

However, it is generally hard to formulate reliable human activity from video data, because of a variety of problems, such as shifting backgrounds, and poor video quality. The two primary issues that are raised by different human activity identification systems are: "Which action is performed?" We also refer to these as "the action recognition task" (also known as, say, final_scores) and "Where exactly in the video? It is also known as the localization problem. Frames are known collections of photographs. The main goal of an action recognition task is to analyze the input video clips to determine what human activities they contained next. They pattern human behaviour and so each human action is different and it's hard to identify. The other difficult problem is to create such deep learning predictive model to predict human behaviour, within benchmark datasets that can be used for assessment. Given the enormous success of the ImageNet [6] dataset for image processing, the action recognition dataset community has published to advance this field of research multiple benchmark datasets. If we look at deep learning model for video data processing it can be treated in the same way that image processing is handled in terms of how much computing power and number of input parameters are required to train the model.

The aim of this research project is to demonstrate what is currently possible in terms of activity recognition using video analysis with deep learning techniques written in python. In this work we aim to outline the most significant human action recognition deep learning models, analyze them and present how current deep learning algorithms used to solve human activity recognition problems in realistic videos take advantage and disadvantages. Our study then based on the quantitative analysis with the recognition accuracies reported by

authors in the literature, will identifies the state of the art deep architectures in action recognition and next will give the current trends and open problems for the future works in this filed. Additionally, it will provide an overview of the state of the art of HAR using wearable and ambient sensors as well as will test the performance of some deep learning approaches, such as deep belief networks or convolutional neural networks over a widely used benchmark dataset for HAR. The subsequent sections will report and elaborate the results and an in depth analysis and discussion of the results, together with experimental results. This study not only provides the potential to validate the practical usefulness of deep learning methodologies for HAR but will also help set the future direction for this emerging area in the digital era.

II. PRELIMINARIES

A human motion recognition analysis using the CNN approach was recently developed using computational intelligence technology by [8]. The authors achieved better accuracy of model, evaluated measuring parameters like low time complexity and reduced execution time of model in the paper. [9] detailedly narrated the three pillars of HAR and covered the period from 2011 to 2021. They further review and present recommendations for a better HAR design, its reliability and stability. Their five major findings were: (1) HAR can be divided into three major pillars (devices, AI, and applications) (2) HAR has dominated the healthcare industry (3) Hybrid AI models are still at their infancy with many needs to do to achieve the stable, reliable design. Additionally, (1) these trained models require solid prediction, high accuracy, generalization, and lastly, achieving the goals of the application in a class free and unbiased manner, (2) little work was seen in abnormality detection, (3) virtually no work was done in actions prediction. We conclude that: Moreover, HAR industry will evolve in terms of the three pillars electronic devices, applications and the type of AI. (b) The harvesting industry in future will be driven more and more by the power of AI.

In [10], they then present a new deep learning based human activity recognition system, which tracks and extract human body in each frame of the video stream, and abstract human silhouettes for creating binary space time maps(BSTMs) to represent human activity in specified time window. Extracting features from BSTMs and classifying activities, they've used convolutional neural network (CNN). To evaluate their approach, they did multiple tests using three public datasets. [11] carried out an update extending previous related surveys, but also focuses on a joint learning framework that identify the temporal and spatial extent of action in videos. Dense trajectories were used as local features to represent the human action. [12] presented a review of the state-of-the-art, showing the overall development of identifying suspicious behavior from surveillance recordings over the past few years. they gave a quick overview of the issues and difficulties associated with recognizing suspicious human activity.

The intelligent human action recognition system, presented in [13], then proposes an automatic feature generation based on image parsing techniques on video image to recognize the human daily activities. Secondly, since processing turns out at a low computational cost and with high accuracy outcomes, the image parsing based approach employed by this work presents great promise [14]. Methods for human activity recognition include manually composed feature based algorithms to state of the art AI based deep learning. Authors such as [15] have surveyed human activity recognition, dividing the study's scope between the data modalities and applications. The study was then further subdivided with respect to different HAR activities using different model development techniques. The analysis of the sampling distribution for the major classification is through the unimodal and multimodal HAR approaches. Unimodal categories are grouped as Space-Time, stochastic, rulebased and shapebased models. A thorough literature evaluation was done [7] for HAR for production and logistics, covering the state of the art HAR methods, statistical pattern recognition and deep architectures in the paper.

The industrial applications of this work are advantageous. Vision based human action recognition was surveyed by [6], who divided the entire study into the following categories: A handcrafted feature and feature learning-based method were used, and the authors described the different techniques and the specifics of how they should be put into practice. The authors also draw attention to relevant material that supports HAR methods at the minute level and is based on categories of human activity, including elementary human activities, gestures, behaviours, interactions, group actions, and events. Similarly, to this, [17] looked at both custom made and learning-based methods for action recognition. The development of cutting-edge activity recognition methodologies in terms of activity representation and HAR classification algorithms was the focus of a review by [18]. The classification of classification approaches is based on template, discriminative, and generative models, while the classification of representation elements is based on global, local, and advanced depth-based. The HAR dataset and the succinctly described models demonstrate performance accuracy in the experimental results. [11] focused on their research on the new trend of integrating all the knowledge acquired so far into a real-life environment. A dataset already published following this philosophy was used to identify the different actions studied. The paper explores new designs and architectures for models inspired by the ones which have yielded the good results in the literature. [19] also explore two deep learning-based approaches, namely single frame Convolutional Neural Networks (CNNs) and convolutional Long Short-Term Memory to recognise human actions from videos. We consider two different deep learning models and their combined formulation to predict action in the video. After applying the models for prediction, determination of which model performs better and select an appropriate model to obtain better efficacy. The proposed models are CNN and LSTM, which are two of the most recent and effective methods for action prediction in videos, as deep learning mechanisms.

III. PROPOSED APPROACH

In this work, we apply deep learning technique to classify human actions from video data. In building the human activity recognition system, we implement two models: Double Frame Convolutional Neural Networks and ConvLSTM. Then what we do is starting with collecting video data and deduce frames from the video files and pre processing them. Next, a datasets is created in which each class has a fixed amount of images. Finally, the model [9] is trained using this dataset.

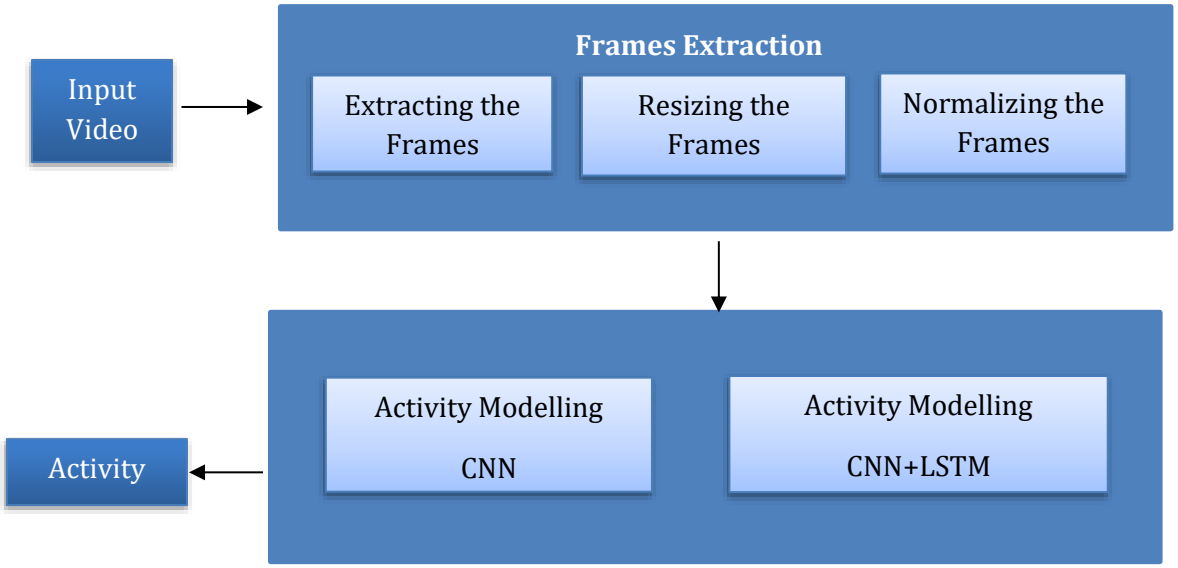

Figure 1. Architecture of the Proposed Method

Before training, the labels (activities) are One Hot Encoded. One way of representing categorical variables as numerical values is to use that of One hot encoding, where every unique value would be represented by a

separate binary feature. [11] For problems such as ours, this technique is especially helpful, as they are multi classification problems. With one hot encoding, we can guarantee that each class is treated identically without other classes assuming some ordinal relationship to classes. Once this is done, we have room to train our models [20]. Later, we will explain more in detail about the two models' architecture that we're using. We will use early stopping callback in training of the model to check the validation loss. When fitting on the model the epochs parameter is set to 50 meaning the total number of times our training dataset can traverse through the network. But if (and when) the validation loss gets no better than 15 epochs (hyperparameter defined), then the early stopping callback will stop training early. When the training stops, the best weights are restored. Early stopping helps reduce the model training time, and also then there will reduce or do not happen overfitting by stopping the epochs before their epoch of overfitting [19].

A. Feature Extraction

The frame extraction pre-processing is one of the prime parts for the performed video applications. Machine learning algorithms based on video can only process it if we have access to the individual frames. In the video frames extraction process, each video file within the dataset's class directories is iterated and the video frames extracted by reading each video file using OpenCV's Video Capture method, frames extracted by iterating through each video frame to resize all the frames to fixed dimensions (64x64 pixels) and normalizing each frame pixels values by dividing it by 255 [13]. Then, the preprocessed frames are pushed in a temporary list for each class. Then we try to extract frames from all of the videos from all action categories and run a function when we want to do it. To make the training easier and in turn to improve the performance and accuracy of the model, we resize the frames to a fixed height and width of the image and normalize the pixel values from 0 to 1.

a. Double Frame CNN

An Architecture of the Convolutional Neural Network (CNN) is described in [19] containing two convolutional layers, a batch normalization layer, a max pooling layer, a global average pooling layer, and two fully connected (dense) layers. We initialize an empty model object that can take successive layers added to the model. The model's first convolutional layer applies a set of 64 (called kernels) filters to the image, each kernel of 3 x 3 size. ReLU is the activation function we use for the layers and is defined as

$$f(x) = \max(0, x)$$

b. Convolutional LSTM

Convolutional LSTM (ConvLSTM) is a variant of the Long Short-Term Memory (LSTM) network architecture that is designed to handle sequential data with spatial structures, such as video data. 2D ConvLSTM is a type of Convolutional LSTM that is designed for processing spatiotemporal data in the form of two-dimensional (2D) arrays, such as image sequences or video frames. Traditional CNNs are good at working with image data, hence are suitable for extracting spatial features from individual frames [20]. On the other hand, LSTM networks are good at modeling temporal dependencies, and hence are suitable for working with sequence data. ConvLSTM combines the power of CNN and LSTM networks to effectively capture both spatial and temporal features of the data, hence making it a suitable approach for solving computer vision problems like video classification [11].

IV. RESULTS AND DISCUSSION

This chapter presents the results obtained from evaluating Human Activity Recognition (HAR) using deep learning models. The analysis includes the comparative performance metrics such as accuracy, precision, recall, and F1 score for each model. Figures 1 to 5 provide a visual representation of these metrics across the Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN),

and combined CNN-LSTM models. Each bar chart highlights the strengths and limitations of each algorithm, allowing for a clear assessment of which models most effectively perform HAR tasks.

Table 1. compiles these performance results, providing a consolidated view of the metrics to support further analysis and interpretation. The findings in this chapter reveal insights into each model's effectiveness in correctly identifying and classifying human activities, thereby guiding the selection of optimal models for HAR applications.

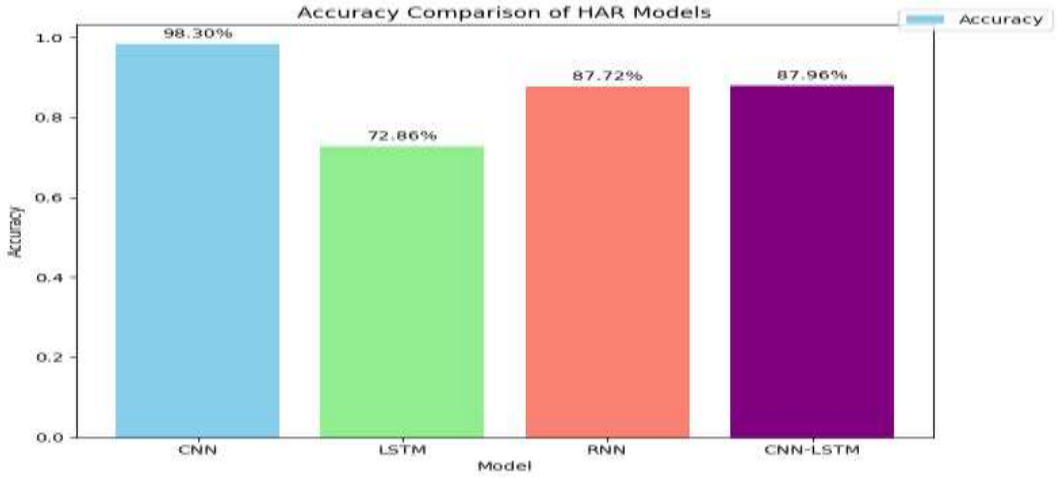

Figure 2. Accuracy Comparison

From Figure 2 above, the CNN algorithm performed highest in terms of Accuracy, followed by the CNN-LSTM model, then the RNN model. It is observed that the LSTM model had the lowest performance in this metric.

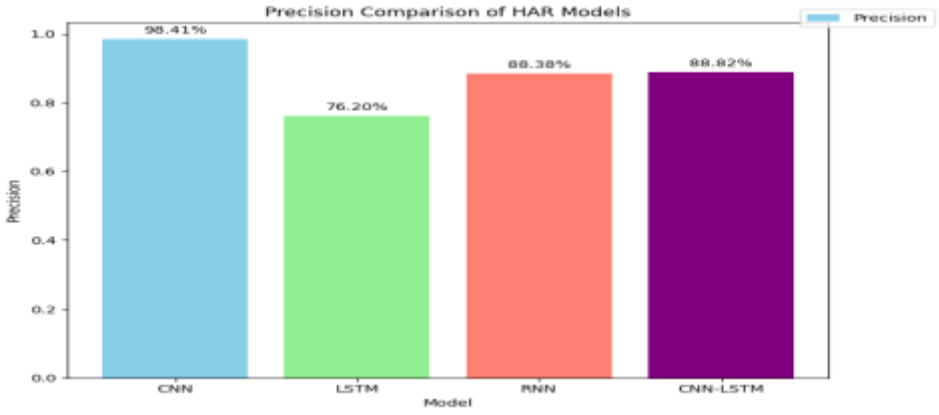

Figure 3. Precision Comparison

From Figure 3 above, the CNN algorithm performed highest in terms of Precision, followed by the CNN-LSTM model, then the RNN model. It is observed that the LSTM model had the lowest performance in this metric.

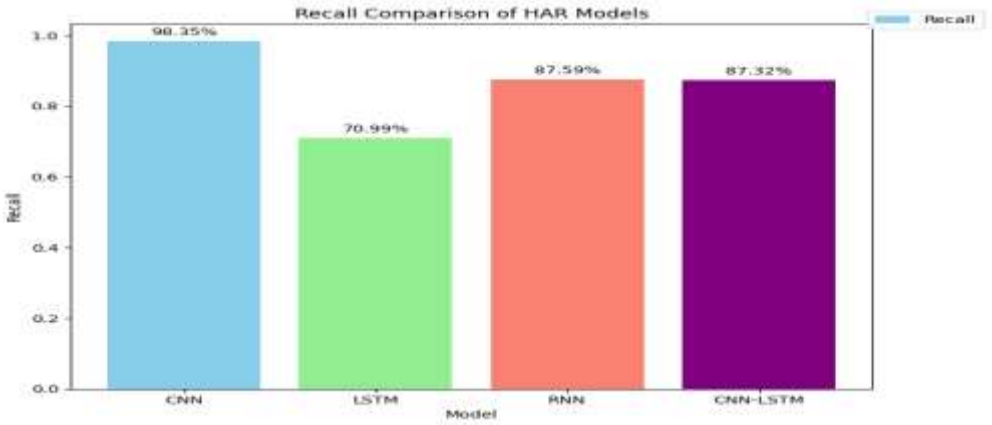

Figure 4. Recall Comparison

From Figure 4 above, the CNN algorithm performed highest in terms of Recall, followed by the CNN-LSTM model, then the RNN model. It is observed that the LSTM model had the lowest performance in this metric.

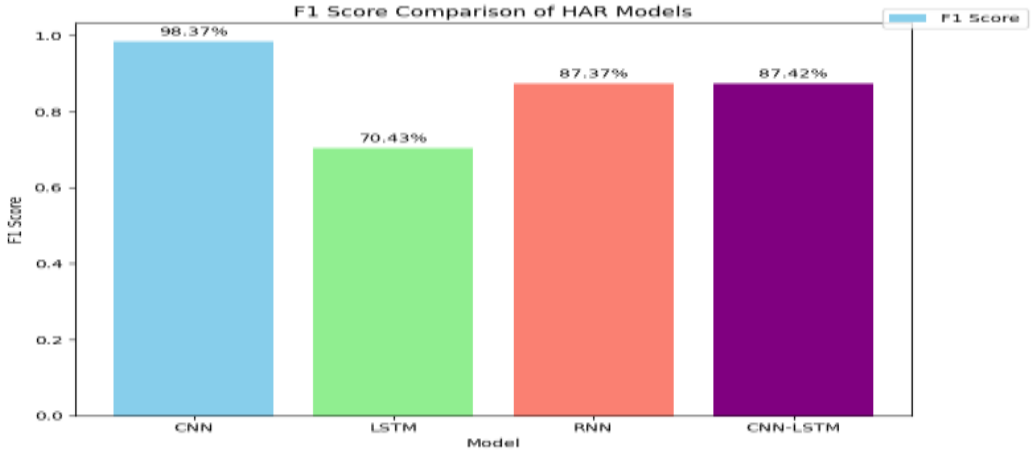

Figure 5. F1 Score Comparison

From Figure 5 above, the CNN algorithm performed highest in terms of F1 Score, followed by the CNN-LSTM model, then the RNN model. It is observed that the LSTM model had the lowest performance in this metric.

Table 1. Summary of Performance Metrics

Model	Accuracy	Precision	Recall	F1 Score
CNN	0.9830	0.9841	0.9835	0.9837
LSTM	0.7286	0.7620	0.7099	0.7043
RNN	0.8772	0.8838	0.8759	0.8737
CNN-LSTM	0.8796	0.8882	0.8732	0.8742

A. Confusion Matrix

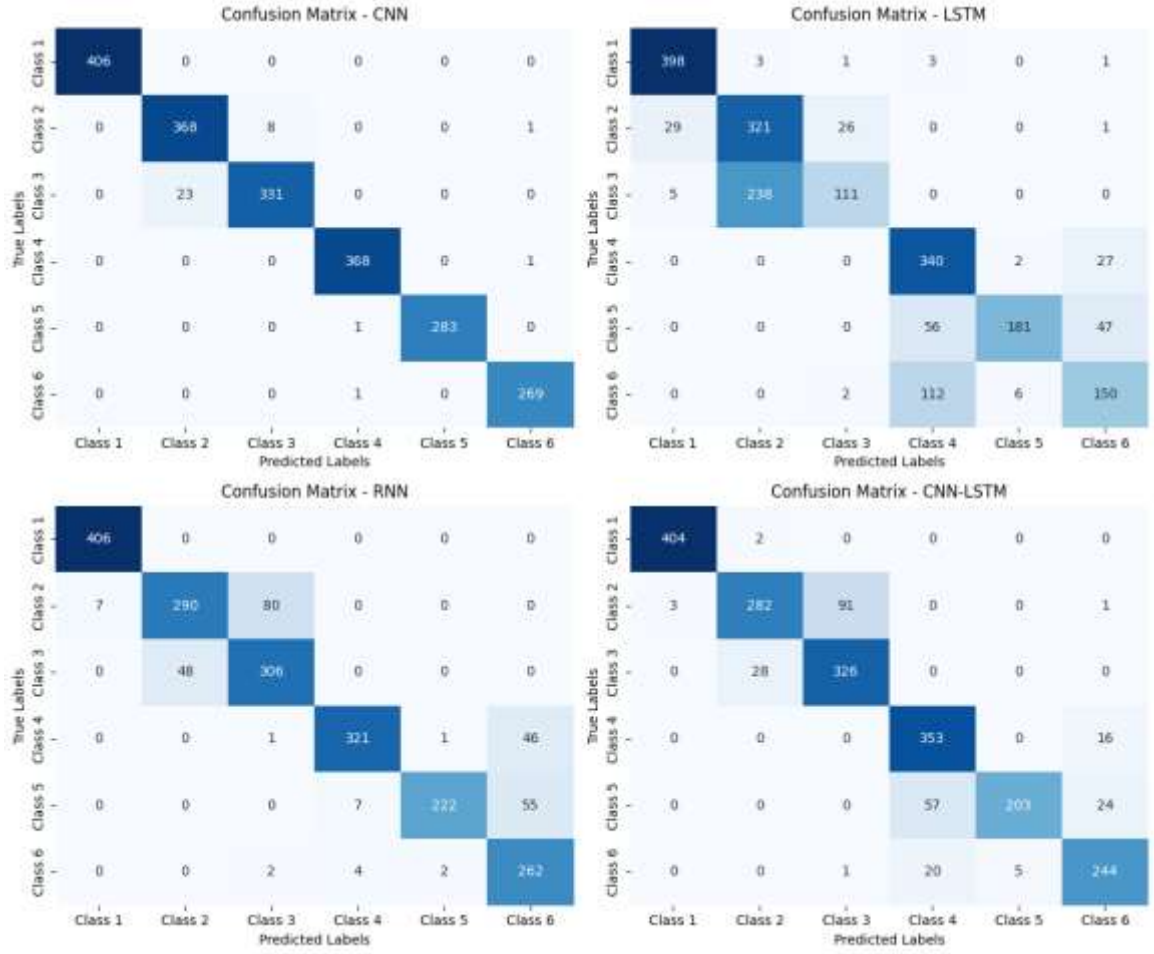

Figure 6. Comparison of Confusion Matrices for CNN, LSTM, RNN, and CNN-LSTM Models

The confusion matrix provides a detailed breakdown of how each model classifies instances across all activity classes. In the CNN confusion matrix, the model demonstrates strong performance, with most predictions correctly aligned along the diagonal. For example, it classifies 406 instances of Class 1 and 368 instances of Class 4 correctly. However, a few misclassifications are present, such as 8 instances of Class 2 being incorrectly classified as Class 3, indicating a slight challenge in distinguishing these specific activities.

The LSTM model has more severe misclassifications, than the regular classification model. It correctly classes 398 instances of Class 1 and 340 of Class 4, but some of the classes give it some difficulties; it misclassifies 26 instances of Class 2 as Class 3 and 47 instances of Class 5 as Class 6. So this pattern indicates that the LSTM model can be misled into thinking two similar activities are different. Likewise, the RNN and CNN-LSTM results are mixed, correctly predicting 406 instances of Class 1 and misclassifying 80 instances of Class 2 as Class 3 and misclassifying 91 instances of Class 2 as Class 3. These numbers shows that CNN performs best in overall, while the other models fall short in similar activity, which may provide a hint of model improvement possibilities.

Class 1: Walking
Class 2: Walking Upstairs
Class 3: Walking Downstairs
Class 4: Sitting

Class 5: Standing
Class 6: Lying Down

B. Discussion of Results

The performance of the Human Activity Recognition (HAR) models CNN, LSTM, RNN, and CNN-LSTM has been evaluated based on key metrics: We can also calculate accuracy, precision, recall, F1 score and confusion matrix results. Performance of activities classification on each of these model is significantly different, with CNN being the most accurate, CNN-LSTM, RNN, and finally LSTM.

a. Accuracy

The CNN model performed best; achieving an accuracy of 98.30%, which shows that most activities can be successfully classified by this model. The high accuracy implies that the features needed for the activity recognition are well learnt and can distinguish by CNN. However, the LSTM model achieved accuracy of 72.86%, which is less compared to the accuracy of other class, and it is unable to recognize some classes well. Intermediary accuracies of 87.72% and 87.96% from RNN (experimental) and CNN's LSTM (experimental) models, respectively, show the RNN and CNN-LSTM models were successful but not CNN in discerning between activities.

b. Precision

Furthermore, CNN has the highest precision of 0.9841, indicating that it performs very few incorrect positive predictions, which is important in applications where one misinterpreted activity will lead to a wrong classification. Similar precision, as RNN, CNN-LSTM has scores of 0.8838 and 0.8882 respectively, which is reasonable but lower than CNN. This LSTM was more prone to more false positives with a precision of 0.7620 than it was to consistently make accurate predictions.

c. Recall

Regarding the recall measure, CNN still obtains the best result, with a score of 0.9835 indicating that CNN can retain most true instances of each activity. Recall is important here as it is very important that the model, in fact, can identify true instances for all activity classes. Recall values of 0.8759 from the RNN and 0.8732 from CNN-LSTM models were achieved, which are reasonable but demonstrate missed instances for some classes. Finally, recall 0.7099 reveals that LSTM model does not recall all or all 0 cases resulting in more false negatives.

d. F1 Score

In fact, CNN's superiority is also reflected in the F1 score, which also has value of 0.9837 (a balanced metric of precision and recall). The high F1 score of this result indicates that CNN is generally robust, and balances precision and recall well. At the end, we used RNN and CNN-LSTM with F1 score of 0.8737, 0.8742 respectively, which indicates reasonable performance but still has some room for improvements in tradeoff between precision and recall. LSTM's F1 score of 0.7043 indicates that LSTM's weakness in efficiently handling both false positives and false negatives is partially justed by its comparison, which is why LSTM failed with this dataset.

e. Confusion Matrix

The confusion matrix results further show how each model classifies each activity correctly and visually displays the results. To circumvent the problem with CNN of misclassifying instances in all classes effectively, we use a confusion matrix from CNN's output and we get the result that CNN does classify most instances and there are less off diagonal values, which means less misclassification. In particular, CNN does very well on activities that are difficult to distinguish between, such as walking upstairs vs walking downstairs, representing only 8 misclassifications from Class 2 to Class 3. However, the confusion matrix of the LSTM model shows that such misclassifications can be found for more than several classes (26 misclassifications

between Class 2 and Class 3, and 47 misclassifications between Class 5 and Class 6), indicating the difficulty to distinguish similar activities. The RNN and CNN-LSTM models have moderate misclassification rate in Class 2 and Class 5 which implies they can capture temporal feature, yet not all of the activities are separated as well as CNN.

However, overall, CNN consistently achieves high scores of accuracy, precision, recall, F1 score, and confusion matrix, while LSTM performs worse on the overall metroplitnicity and most in activities at which has characteristics overlap.

V. CONCLUSION

The Human Activity Recognition (HAR) study, using CNN, LSTM, RNN, and CNN–LSTM, demonstrates an outstanding performance for CNN as the best performing model. In particular, CNN demonstrates the strongest capability to capture essential spatial features for effective classification between activities, leading to higher accuracy, precision, recall, and F1 score compared to other classifiers. Since high classification accuracy is critical for HAR tasks, CNN turns out to be more suitable as compared to other networks, which makes CNN particularly suitable for HAR tasks. Moderate performance was achieved by the RNN and CNN-LSTM models, which demonstrate the ability to learn temporal patterns whilst still making some classification errors between similar activities. Performance metrics for the LSTM model were the lowest, as the LSTM model did not perform well on classes where characteristics overlap.

Conflicts of Interest

I declare that there are no conflicts of interest regarding the publication of this research. No sources of funding, affiliations, and any relevant financial or non-financial relationships have influenced this research or its outcomes.

Data Availability Statement

All sources of data and information used for this research have been duely acknowlewdged and a list of reference provided

Acknowledgements

The Authors of this manuscript wish to acknowledge the the entire university management and and tetfund department for the enabling environment to undergo this research. The various resources and data sources are hereby acknowledged and a list of references provided.

VI. REFERENCES

1. Chuanwei Zhou, Chunyan Xu, and Zhen Cui, “Progressive Bayesian Inference for Scribble-Supervised Semantic Segmentation,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, pp. 3751-3759, 2023. Google Scholar | Publisher Link
2. Christoph Angermann et al., “Correction: Unsupervised Single-Shot Depth Estimation Using Perceptual Reconstruction,” *Machine Vision and Applications*, vol. 34, 2023. Google Scholar | Publisher Link
3. Angelamaria Cardone et al., “Collocation Methods for Volterra Integral and Integro-Differential Equations: A Review,” *Axioms*, vol. 7, no. 3, pp. 1-19, 2018. Google Scholar | Publisher Link
4. Bhushan Marutirao Nanche, Hiren Jayantilal Dand, and Bhagyashree Tingare, “Human Activity Recognition Using Deep Learning: A Survey,” *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 9, no. 3, pp. 605-610, 2020. Google Scholar | Publisher Link
5. Stephen Mkegh Nengem, “Symmetric Kernel-Based Approach for Elliptic Partial Differential Equation,” *Journal of Data Science and Intelligent Systems*, vol. 1, no. 2, pp. 99-104, 2023. Google

Scholar | Publisher Link

6. Djamila Romaissa Beddiar et al., "Vision-Based Human Activity Recognition: A Survey," *Multimedia Tools and Applications*, vol. 79, pp. 30509-30555, 2020. Google Scholar | Publisher Link
7. Christopher Reining et al., "Human Activity Recognition for Production and Logistics-A Systematic Literature Review," *Information*, vol. 10, no. 8, pp. 1-28, 2019. Google Scholar | Publisher Link
8. Muhammad Ahmed Raza, and Robert B. Fisher, "Vision-Based Approach to Assess Performance Levels While Eating," *Machine Vision and Applications*, vol. 34, pp. 1-14, 2023. Google Scholar | Publisher Link
9. Kailong Liu et al., "A Data-Driven Approach With Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3170-3180, 2021. Google Scholar | Publisher Link
10. David J. Stracuzzi et al., "Data-Driven Uncertainty Quantification for Multisensor Analytics," *Proceedings, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent*, vol. 10635, 2018. Google Scholar | Publisher Link
11. Daniel Garcia-Gonzalez et al., "Deep Learning Models for Real-Life Human Activity Recognition from Smartphone Sensor Data," *Internet of Things*, vol. 24, pp. 1-22, 2023. Google Scholar | Publisher Link
12. Hjortur Bjornsson, and Sigurdur Hafstein, "Advanced Algorithm for Interpolation with Wendland Functions," *Informatics in Control, Automation and Robotics*, *Lecture Notes in Electrical Engineering*, vol. 720, pp. 99-117, 2020. Google Scholar | Publisher Link
13. Y.C. Hon, R. Schaback, and X. Zhou, "An Adaptive Greedy Algorithm for Solving Large RBF Collocation Problems," *Numerical Algorithms*,vol. 32, pp. 13-25, 2003. Google Scholar | Publisher Link
14. Juan Carlos Niebles, Hongcheng Wang, and Li Fei-Fei, "Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words," *International Journal of Computer Vision*, vol. 79, pp. 299-318, 2008. Google Scholar | Publisher Link
15. Md Shariful Islam et al., "Approximate Solution of Systems of Volterra Integral Equations of Second Kind by Adomian Decomposition Method," *Dhaka University Journal of Science*, vol. 63, no. 8, pp. 15–18, 2015. Google Scholar | Publisher Link
16. Michalis Vrigkas, Christophoros Nikou, and Ioannis A. Kakadiaris, "A Review of Human Activity Recognition Methods," *Frontiers in Robotics and AI*, vol. 2, pp. 1-28, 2015. Google Scholar | Publisher Link
17. Fan Zhu et al., "From Handcrafted to Learned Representations for Human Action Recognition: A Survey," *Image and Vision Computing*,vol. 55, no. 2, pp. 42-52, 2016. Google Scholar | Publisher Link
18. Shugang Zhang et al., "A Review on Human Activity Recognition Using Vision-Based Method," *Journal of Healthcare Engineering*, vol. 2017, no. 1, pp. 1-31, 2017. Google Scholar | Publisher Link
19. Sheryl Mathew et al., "Human Activity Recognition Using Deep Learning Approaches and Single Frame CNN and Convolutional LSTM," *Arxiv*, pp. 1-16, 2022. Google Scholar | Publisher Link
20. Palle Jorgensen, and Feng Tian, "Discrete Reproducing Kernel Hilbert Spaces: Sampling and Distribution of Dirac-Masses," *Journal of Machine Learning Research*, vol. 16, no. 96, pp. 3079−3114, 2015. Google Scholar | Publisher Link