Human Activity Recognition Using Feed Forwarding Neural Network

Roopa Devi E M
Department of Information Technnology
Kongu Engineering College Erode,India

Abstract:

Human Activity Recognition (HAR) is a transformative technology that enables systems to understand and interpret the actions of people in real-time. By recognizing patterns in human behavior, HAR has the potential to revolutionize industries such as healthcare, security, and smart homes. This study explores HAR using deep learning models, with a focus on Convolutional Neural Networks (CNN) and Deep Convolutional Neural Networks (DCNN) as methodologies. CNN and DCNN are adept at capturing spatial hierarchies and complex data relationships, enabling accurate recognition of human activities. The research delves into data preprocessing, model design, training, and evaluation, showcasing the effectiveness of CNN and DCNN in HAR tasks. Additionally, within the feedforward neural network framework, ResNet 34 and ResNet 50 architectures are investigated, leveraging residual connections to address challenges like vanishing gradients. Comparative analysis reveals ResNet 50 as the standout performer, achieving an impressive in recognizing human activities. The potential of advanced deep learning techniques, particularly ResNet 50, in enhancing HAR systems across diverse applications.

IIntroduction:

Human Action Recognition (HAR) using Machine Learning represents a captivating domain within computer vision, aiming to educate machines on comprehending and interpreting human actions from visual data. This field boasts a myriad of applications, spanning surveillance, robotics, healthcare, and human-computer interaction. The primary goal is to devise algorithms and models capable of automatically discerning and categorizing human actions depicted in images or videos. The automatic recognition and understanding of human actions from visual data significantly enhance the capabilities of intelligent systems. In surveillance, HAR facilitates the detection of anomalous activities, bolstering security measures. In human-computer interaction, it enables more natural interfaces, allowing systems to intelligently respond to users' gestures or movements. Sports analytics benefits from HAR for performance analysis and strategy optimization, while in healthcare, it plays a pivotal role in monitoring and assessing physical activities for rehabilitation or wellness applications. Over time, the evolution of HAR methodologies has seen a shift towards the integration of deep learning techniques. Feedforward neural networks, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks, have demonstrated significant success in addressing the intricate nature of human actions. CNNs excel in spatial feature extraction, capturing crucial information about body parts and their spatial relationships, making them highly effective in recognizing complex human actions from visual data. Utilizing Feedforward Neural Networks, particularly CNNs, for Human Action Recognition (HAR) signifies an advanced approach within computer vision, aiming to automatically identify and categorize human actions from visual data, such as video frames. CNNs, originally designed for image classification, exhibit remarkable efficacy in capturing spatial features and patterns crucial for recognizing intricate human actions. HAR using CNNs represents an intelligent and automated solution in computer vision for recognizing complex human actions in video data. The adaptability and effectiveness of CNN architectures, combined with ongoing advancements, position HAR as a critical technology applicable across diverse domains, including surveillance and healthcare. In Human Activity Recognition (HAR), Feedforward Neural Networks are employed in the form of Long Short-Term Memory (LSTM) networks, a specialized type of Recurrent Neural Network (RNN). LSTMs are particularly effective for sequential tasks, as they excel at capturing temporal dependencies in action sequences, making them highly suitable for tasks involving video data. Alternatively, Feedforward Neural Networks like ResNet are utilized, leveraging residual connections to facilitate the training of exceptionally deep networks.This section thoroughly explores ResNet architecture, elucidating the role of residual blocks and skip connections in overcoming challenges related to vanishing gradients. Understanding these principles is essential for adapting ResNet to the temporal nature of sensor data in HAR. The methodology encompasses training ResNet models on labeled sensor data, covering activities such as walking, running, and various daily-life movements. The paper details preprocessing steps, model hyperparameters, and the incorporation of temporal dependencies within ResNet, with special attention given to fine-tuning for optimizing the network for HAR tasks.

II Literature Review:

Qiao-Yuan Yao, Po-Lin Chen, and Tzung-Shi Chen [1] introduced a home care assistance system aimed at elder human activity recognition (HAR), integrating 2-D LiDAR technology and deep learning approaches. The system employs 2-D LiDAR to capture spatial data from a room's interior, utilizing clustering algorithms to identify high-density regions likely to represent objects. By distinguishing between human and nonhuman clusters, the system records the coordinates of human clusters to create trajectory graphs. These graphs are analyzed using spatial-temporal graph convolutional networks (ST-GCN) to classify human activities with precision. Additionally, the system is designed to recognize unusual trajectory patterns, such as those indicating falls, and promptly alerts caregivers with an emergency signal. The study also details the development of a user-friendly interface for the assistance system, emphasizing both safety and privacy for users. By leveraging the unique attributes of LiDAR point cloud data and employing the ST-GCN model, the proposed approach demonstrates strong classification performance and reliable accuracy.

Shady Younan, in collaboration with Mervat Abu-Elkheir [2], presents an extensive review of the advancements in machine learning and deep learning algorithms for Human Activity Recognition (HAR) using sensor data. The study examines how various sensor types and modalities influence algorithm selection, advocating for the use of generic sensor types to enhance integration and user experience. Additionally, it highlights the challenges associated with the complexity and deployment of deep neural networks, proposing solutions such as scaling down architectures or utilizing optimized versions for edge and mobile devices. Incremental learning is discussed as a viable alternative to retraining models from the ground up, with strategies to reduce annotation requirements and identify meaningful unlabeleddata. Furthermore, the paper explores the concept of human-AI collaboration in model updates, including approaches like Learn, teacher-student self-training, and multitask self-supervision methods employed in SelfHAR.

Wei Zhong Tee, Rushit Dave, Jim Seliya, and Mounika Vanamala [3] provide a comprehensive review of advanced human activity recognition (HAR) models built on deep learning architectures, including those employing Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid architectures combining both. HAR approaches are categorized into vision-based recognition, which relies on video data from cameras, and sensor-based recognition, which utilizes data from mobile or wearable sensors. Frequently used datasets in HAR research include OPPORTUNITY, Skoda Checkpoint, UCI-HAR, WISDM, MHEALTH, and PAMAP2. The paper emphasizes the efficiency and capability of deep learning methods over traditional machine learning algorithms such as Support Vector Machines (SVM), k-Nearest Neighbor (kNN), and Random Forests (RF). The reviewed models leverage CNNs, LSTMs, and hybrid architectures, demonstrating their ability to handle diverse tasks. Challenges in HAR systems often stem from data preprocessing and feature extraction, rather than the classifiers themselves. The study also highlights model performance, showcasing examples like a unidirectional LSTM-based DRNN achieving 97.8% overall accuracy and 97.4% average precision on the USC-HAD dataset, as well as a CNN model with bidirectional LSTM layers demonstrating high accuracy across multiple datasets.

Saurabh Gupta [4] presents a literature review discussing various methodologies for human activity recognition (HAR) utilizing wearable sensors, machine learning algorithms, and hybrid deep learning frameworks. The review categorizes machine learning algorithms for HAR into three primary types: discriminative, generative, and hybrid. It examines the strengths and limitations of different techniques, including decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN). The study contributes by applying HAR to classify complex human actions using the "WISDM

Smartphone Smartwatch Activity and Biometrics Dataset," emphasizing the potential of smart devices equipped with integrated sensors. The proposed hybrid deep neural network model, CNN-GRU, is evaluated against other advanced models like Inception Time and DeepConvLSTM, demonstrating comparatively better accuracy.

Young Ghyu Sun, Soo Hyun Kim, Seongwoo Lee, Joonho Seon, SangWoon Lee, Cheong Ghil Kim, and Jin Young Kim [5] propose an end-to-end HAR model built on convolutional long short-term memory (LSTM) architecture, leveraging deep learning techniques. The model's performance is assessed using metrics such as Precision, Recall, F1-score, and Matthew's Correlation Coefficient (MCC). Simulation results reveal that the convolutional LSTM model surpasses traditional models in classification accuracy. When compared to an end-to-end model utilizing 3D-CNN, the convolutional LSTM demonstrates superior performance. The model is applicable in various domains, including smart signage, smart homes, and smart healthcare systems.

Haixia Bi, Miquel Perello-Nieto, Raul Santos-Rodriguez, and Peter Flach [6] explore the challenges associated with activity recognition, including the scarcity of annotated training data and the variability in how individuals perform activities. To address these issues, the study proposes a dynamic active learning-based approach. This method is emphasized as an effective strategy for minimizing annotation costs while enhancing model performance in activity recognition tasks. Additionally, the paper introduces the application of a one-class support vector machine (OCSVM) for novelty detection, which facilitates the identification of new activities and patterns within the data. The proposed method, called OCluDAL, leverages novelty detection and clustering techniques to understand the data distribution and selects informative samples based on uncertainty, diversity, and representativeness. Experimental results demonstrate that OCluDAL reduces annotation costs and achieves better performance compared to other methods.

Youngwook Kim, Senior Member, IEEE, and Taesup Moon, Member[7] research efforts have focused on distinguishing targets without time-dependent information and have used spectral analysis of micro-Dopplers.Time-varying signatures extracted from spectrograms have been used to recognize various human activities. Empirical mode decomposition has been successful in recognizing target types. Principal component analysis and linear discriminant analysis have been used to extract feature vectors. The linear predictive code has been suggested for real-time processing. The paper proposes an alternative deep learning approach using DCNNs to overcome the limitations of previous methods. Deep learning has gained interest in machine learning and has achieved excellent performance in various domains. DCNN is a successful deep learning algorithm that uses convolution filters, nonlinear activation functions, and pooling to extract features and classify data. The paper uses a 2-s spectrogram as input to the DCNN and interprets the classification as an image recognition problem. The DCNN in the paper consists of two convolution layers, each with four filters of size 5 x 5, and one fully connected layer.

III Proposed Methodology:

This research focuses on multi-label human activity classification, using a mix of CNN, DCNN, Resnet 34 and primarily Resnet 50 models. The dataset features 15 different classes of Human Activities, classifying 15 different activities. The main goal is to build a predictive model that assigns probability scores to each type of activity in individual activities.

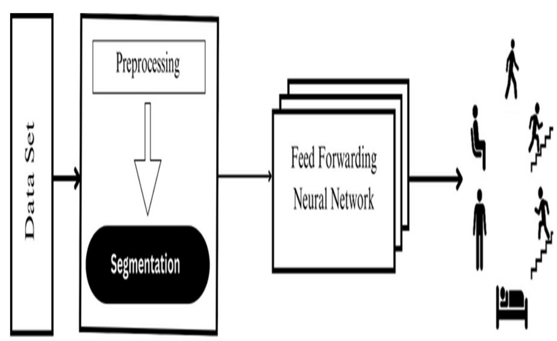

Figure 1: Proposed Framework for Human Activity Recognition

3.1 Data Preprocessing:

The data preprocessing pipeline is a crucial component in enhancing the quality and compatibility of the input data for the Human Activity Recognition (HAR) model. This involves a series of steps to ensure uniformity, cleanliness, and optimal utilization of the available data.

3.1.1 Image Preprocessing:

Resize Images: All images are resized to a consistent dimension of [224, 244] pixels. Standardizing image dimensions facilitates efficient processing and model training.

Normalize Pixel Values: Pixel values are normalized to a standard range. This involves multiplying pixel values by pre-defined factors to transform images from the [0, 255] range to [0, 1].

3.1.2 Dataset Splitting:

Training-Validation Split: The dataset is split into two subsets: 85% is used for training, while 15% is reserved for validation. This partitioning provides a designated portion of the data for evaluating the model's performance.

3.1.3 Label Encoding:

Label Mapping: Human activity labels are encoded into numerical values using a mapping scheme. This allows the model to understand and learn from the categorical labels during training.

3.1.4 Data Loading:

Custom Dataset Class: A custom dataset class is implemented to load file paths, labels, and transformations. This class ensures efficient handling of data during training and validation.

Data Loader Configuration: Data Loaders are configured for both training and validation sets, specifying batch sizes, shuffling, and drop-last functionalities.

3.2 Classification and CNN Models:

CNN:

Human activity recognition using CNN involve a systematic approach. Initiate the process by curating a diverse dataset annotated with human activities. Apply data preprocessing techniques, including resizing and normalization, to ensure uniform input. Design a CNN architecture tailored to capture spatial hierarchies in the data, incorporating convolutional and pooling layers. Implement transfer learning by utilizing pre-trained CNN models such as VGG or ResNet, fine-tuned on the human activity dataset to leverage learned features. Augment the training dataset with variations to enhance model robustness. Perform comprehensive hyperparameter tuning, optimizing learning rates and batch sizes. Divide the dataset into training and testing subsets to ensure an objective evaluation of the model's performance. Use the training set to train the CNN and the testing set to validate its performance, employing metrics such as accuracy, precision, and recall for assessment. Iterate and refine the model based on evaluation results, ensuring it generalizes well to diverse human activities. The proposed CNN methodologies aim for an effective and accurate human activity recognition system with applications in health monitoring, surveillance, and beyond.

DCNN:

Deep Convolutional Neural Networks (DCNNs) are utilized for human detection and activity classification using micro-Doppler signatures derived from Doppler radar data. Unlike traditional supervised learning approaches that depend on manually designed features, DCNNs operate directly on raw micro-Doppler spectrograms. For human detection, the DCNN distinguishes humans from other targets (e.g., dogs, horses, and cars). In human activity classification, the model identifies activities, such as running and crawling. The input to the DCNN is the spectrogram itself, eliminating the need for explicit domain knowledge in feature extraction. These methodologies highlight the potential of deep learning in radar signal processing, applicable in defence, surveillance, and beyond. Future research directions include visualizing higher-layer representations in DCNNs to enhance feature insights. Additionally, careful consideration of computational complexity is crucial for real-time processing applications.

Feed Forwarding Neutral Network:

A feedforward neural network integrating ResNet 34 and ResNet 50 architectures employs residual connections to address the vanishing gradient problem, facilitating training of deeper models. ResNet 34 comprises 34 layers with repeated convolutional blocks, while ResNet 50 adds bottleneck layers for efficiency. These architectures excel in computer vision tasks like image classification.

Resnet 34:

ResNet-34 for human activity recognition involves a systematic approach. The process begins with the compilation of a diverse dataset, meticulously annotated with human activities. Data preprocessing follows, incorporating normalization and standardization techniques. The ResNet-34 architecture, featuring residual blocks for efficient feature extraction, is implemented. Initialization of the network involves pre-trained weights from datasets like ImageNet, with subsequent fine-tuning on the human activity dataset. Training data augmentation, involving various transformations, is coupled with regularization techniques such as dropout. Comprehensive hyperparameter tuning is conducted to optimize learning rates and batch sizes. Iterative refinement of the model, considering potential architecture adjustments, is guided by evaluation outcomes. Ultimately, the optimized ResNet-34 model is deployed for real-time human activity recognition, leveraging its capabilities for robust and effective activity classification.

Resnet 50:

We initiate the process by compiling a diverse dataset annotated with human activities, employing the provided CSV file for labeling. A meticulous 85-15% split ensures robust training and validation sets. Data preprocessing involves normalization and standardization to maintain consistent quality. We implement ResNet-50 and ResNet-34 architectures, fine-tuning with pre-trained ImageNet weights. Data augmentation and regularization techniques, including dropout, enhance model generalization.

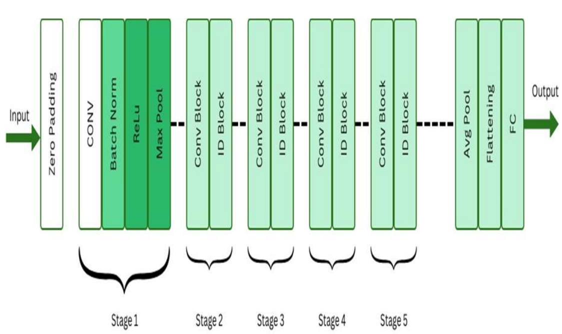

Figure 2: ResNet 50 Model Architecture

Training occurs using a Data Loader with a batch size of 128, utilizing CrossEntropyLoss. Model evaluation on a distinct validation set focuses on accuracy as the primary metric. Comparative analysis involves assessing ResNet-50 and ResNet-34 performance, considering computational complexity. Iterative refinement responds to evaluation outcomes with potential architecture adjustments.

Deployment of optimized models allows real-time human activity recognition. A thorough comparative evaluation emphasizes strengths and weaknesses, utilizing final training and validation accuracies. The conclusion selects the superior model, substantiating the choice with a detailed analysis. This methodology ensures a systematic exploration of ResNet-50 and ResNet-34, facilitating an informed decision on the more effective model within a human activity recognition context.

IV Experimental Result and Analysis:

The dataset contains approximately 12600 images representing various tasks and pixels. The images could represent comparable sets of tasks in various contexts. The dataset was split into two parts: training and testing datasets. The 80 percentage of the data can be considered for training dataset and 20 percentage of the data can be considered for testing dataset.

4.1 Evaluation Metrics:

The performance of machine learning models is assessed using several quantitative evaluation metrics: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). TP refers to accurately predicted positive cases, FP represents incorrectly predicted positives, TN corresponds to correctly identified negatives, and FN indicates incorrectly classified negatives.

Accuracy:

Accuracy is a pivotal for assessing ResNet-50-based human activity recognition models, measures the ratio of correct predictions, evaluating overall performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision: is a crucial in ResNet-50-based activity recognition, gauges accurate positive predictions, with a score of 1 denoting perfect identification of positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

Recall:

Recall is a vital in ResNet-50-based activity recognition, measures the model's ability to identify positive instances accurately, calculated as true positives over actual positives.:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

F1 Score:

The F1 score is a crucial in ResNet-50-based activity recognition, combines precision and recall, offering a balanced assessment of model performance.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recal} \quad (4)$$

Models	Accuracy	Precision	Recall	F1 Score
CNN	0.89	0.90	0.81	0.82
DCNN	0.91	0.89	0.83	0.88
Resnet 34	0.93	0.86	0.87	0.87
Resnet 50	0.96	0.97	0.93	0.94

Table 1: Performance comparison results of various deep learning models

When the above-mentioned algorithm is executed, a respective output is being generative. The output generally deals with displaying the activity that is portrayed in the image. And an example is displayed below

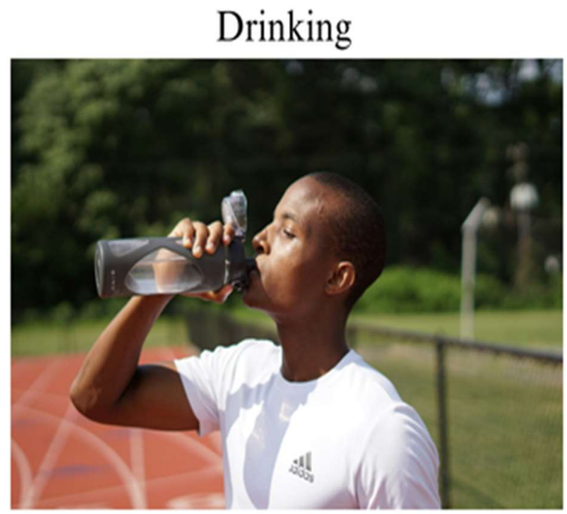

Figure 3: Example ofRecognised Movement

Training Accuracy: Represents the proportion of correctly predicted instances on the training dataset, indicating how well the model performs on seen data. Validation Accuracy: Reflects the accuracy achieved on a separate validation dataset, providing insights into the model's generalization ability on unseen data.

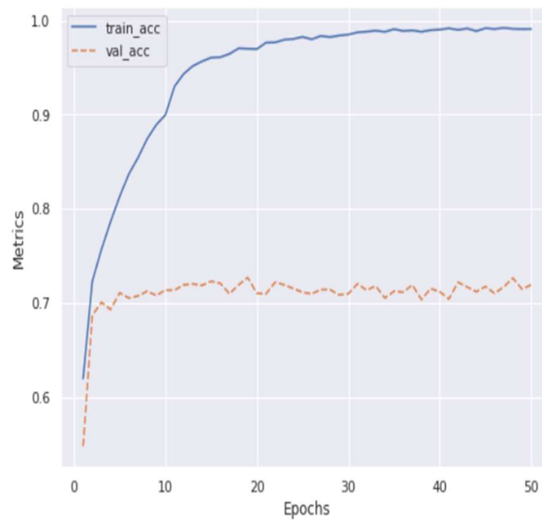

Figure 4: Training and Validation Accuracy Analysis for Human Activity Recognition

The training accuracy typically increases over epochs, demonstrating that the model is improving its predictive performance on the training data. Validation accuracy may initially increase but could plateau or fluctuate as the model learns, indicating its ability to generalize to unseen data. Consistent high training and validation accuracy suggest a well-performing model with good generalization capabilities.

The comparative results of Human Activity Recognition among CNN, DCNN, Resnet 34, and Resnet 50 yields valuable insights. CNN, achieving 89% accuracy, offers simplicity but might struggle with subtle activity recognition. DCNN, with an 91% accuracy rate, shows better performance, indicating a deeper understanding of intricate data relationships. Resnet, stands out with an impressive 93% accuracy, highlighting it with better result. However, Resnet 50 that stands as the star performer with an impressive accuracy of 95%, indicating its unparalleled ability to understand and recognition the activity of the humans accurately.

V Conclusion:

In conclusion, our research highlights the effectiveness of deep learning models, especially ResNet 50, in accurately recognizing human activities from visual data. These results emphasize the potential of cutting-edge deep learning techniques to advance Human Activity Recognition (HAR) systems in various fields, including surveillance and healthcare. Future studies could explore the integration of hybrid architectures and the use of novel sensor technologies to further enhance HAR performance and broaden its real-world applicability.

References:

1. Qiao-Yuan Yao , Po-Lin Chen, and Tzung-Shi Chen,"Human Activity Recognition Using 2-D LiDAR and Deep Learning Technology",25 July 2023; accepted 13 September 2023. Date of publication 19 September 2023; date of current version 27 September 2023.

2. Shady Younan and Mervat Abu-Elkheir , Member, IEEE "Deep Incremental Learning for Personalized Human Activity Recognition on Edge Devices",Accessed:SUMMER 2022 IEEE Canadian Journal of Electrical And Computer Engineering, VOL. 45, NO. 3, SUMMER 2022.

3. Wei Zhong Tee, Rushit Dave, Jim Seliya, Mounika Vanamala," A Close Look into Human Activity Recognition Models using Deep Learning",Acessed:2022.

4. Saurabh Gupta," Deep learning based human activity recognition (HAR) using wearable sensor data",Acessed: International Journal of Information Management Data Insights 1 (2021).
5. Young Ghyu Sun1, Soo Hyun Kim1, Seongwoo Lee1,Joonho Seon1, SangWoon Lee2, Cheong Ghil Kim3 and Jin Young Kim," Performance of End-to-end Model. Based on Convolutional LSTM for Human Activity Recognition", Received 09 February 2022; Accepted 31 May 2022;Publication 27 August 2022.

6. Youngwook Kim, Senior Member, IEEE, and Taesup Moon, Member, IEEE,"Human Detection and Activity Classification Based on Micro-Doppler Signatures Using DeepConvolutional Neural Networks",Accessed:VOL. 13, NO. 1, JANUARY 2016.

7. Haixia Bi , Miquel Perello-Nieto, Raul Santos-Rodriguez, and Peter Flach,"Human Activity Recognition Based on Dynamic Active Learning",Acessed: IEEE Journal of Biomedical And Health Informatics, VOL. 25, NO. 4, APRIL 2021

8. Haixia Bi , Miquel Perello-Nieto, Raul Santos-Rodriguez, and Peter Flach," Human Activity Recognition Based on Dynamic Active Learning",Acessed: IEEE Journal of Biomedical And Health Informatics, VOL. 25, NO. 4, APRIL 2021