IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 3, May-June 2025

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

94-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203051

Human Activity Recognition Using Advanced Deep Learning Approaches

Parthasarathy G

II M.Sc.-CS, Dept. of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

B. Suresh

Asst. Professor, Dept. of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

ABSTRACT: Human Activity Recognition (HAR) using image data has become an important research area, especially in applications such as smart surveillance, fitness tracking, and behaviour analysis. This paper proposes an image-based deep learning framework that combines Self-Supervised Learning (SSL) for feature extraction, Transformer-based sequence modelling, convolutional neural networks (CNNs) for training, transformer-based architectures and Generative Adversarial Networks (GANs) for data augmentation.Self-Supervised Learning (SSL) with Convolutional Neural Networks (CNNs) to effectively classify human activities from static images. The SSL module enhances feature learning by training the model to reconstruct masked regions of input images. The learned features are then used to classify activities using a CNN classifier. The proposed approach achieves high accuracy on a labelled image dataset of human activities. Evaluation results and visual analysis demonstrate the model's strong generalization capabilities in classifying diverse human actions.

KEYWORDS: Human Activity Recognition, Deep Learning, Convolutional Neural Networks, Self-Supervised Learning, Image Classification.

I. INTRODUCTION

Human Activity Recognition (HAR) is a rapidly growing field in computer vision and artificial intelligence, offering impactful applications in areas such as healthcare monitoring, security surveillance, smart homes, and interactive systems. Traditional HAR systems primarily rely on wearable sensors or video sequences to detect and classify human actions. However, these methods often face limitations including high cost, intrusiveness, and complexity in data collection. To overcome these issues, image-based HAR has emerged as a viable alternative that leverages still images to recognize human activities.

In this research, we propose a novel framework that utilizes advanced deep learning techniques—particularly Self-Supervised Learning (SSL) and Convolutional Neural Networks (CNNs)—to accurately classify human actions from images. Unlike sensor-based approaches, our model is non-intrusive and scalable, requiring only labelled and unlabelled image datasets for training and evaluation. The proposed system enhances the feature extraction capability of CNNs by integrating an SSL-based pretraining stage, which learns meaningful representations without the need for manual annotations.

II.. LITERATURE REVIEW

Several studies have focused on HAR using both traditional machine learning and deep learning techniques. Early work in this domain employed handcrafted features such as Histogram of Oriented Gradients (HOG), SIFT, and optical flow to represent human movements. However, these methods are often limited in their ability to generalize across diverse environments.

Gkioxari et al. (2015) introduced an action recognition system based on region-based CNNs that analysis spatial cues from still images. Their work demonstrated that deep learning models could outperform traditional methods in complex action recognition tasks. More recently, self-supervised learning has been applied to HAR to improve model performance in scenarios with limited labelled data. Dyer et al. (2020) presented a framework where models learn invariant representations through transformations, showing significant improvement in downstream classification tasks.

94-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

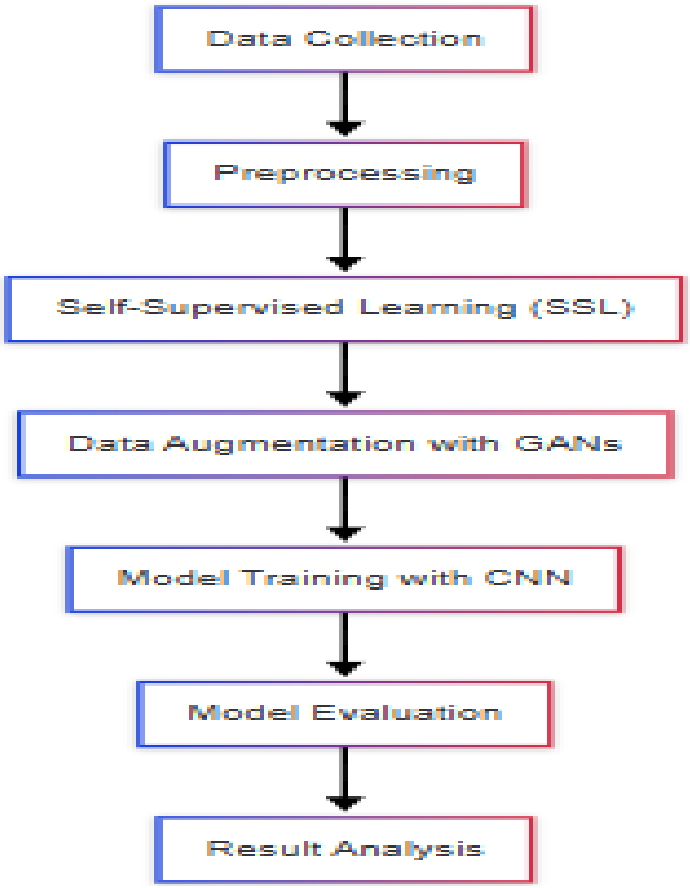
|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203051

These studies highlight the potential of combining CNN architectures with SSL techniques to build robust image-based HAR systems.

III. METHODOLOGY

This section details the comprehensive methodology adopted to develop and evaluate a deep learning-based Human Activity Recognition (HAR) system using image datasets. The approach involves multiple stages, including data preprocessing, self-supervised learning for representation learning, CNN-based classification, model training, and evaluation.



3.1 Dataset Acquisition and Preprocessing

The project utilizes publicly available human activity image datasets. Each image represents a specific physical activity such as walking, running, jumping, sitting, or exercising. To ensure consistency and improve learning, all images are resized to a fixed dimension (e.g., 224×224 pixels), normalized, and augmented for better generalization.

Data preprocessing involved the following steps:

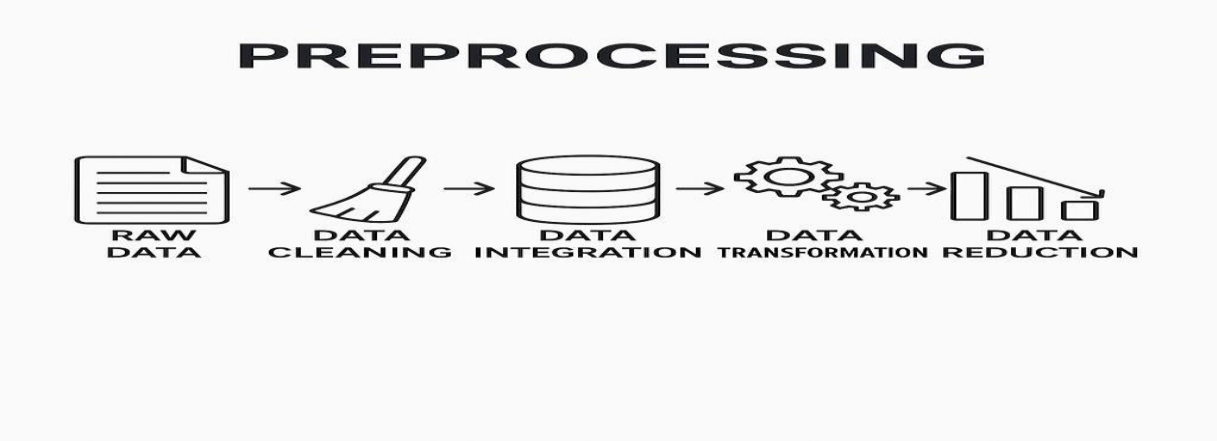
- **Resizing:** All input images were resized to standard dimensions suitable for CNN input layers.
- **Normalization:** Pixel values were normalized to a range of [0, 1] to ensure uniformity in feature distribution.
- **Augmentation:** Techniques such as rotation, horizontal flipping, zoom, and brightness alteration were applied to artificially increase the dataset size and reduce overfitting.Additionally, the dataset was split into training (70%), validation (15%), and test (15%) subsets.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

94-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203051



3.2 Self-Supervised Learning for Feature Representation

Given the limited availability of labeled data, a self-supervised learning (SSL) strategy was adopted as a pretraining step. The goal of SSL is to enable the model to learn meaningful visual representations from the data itself without relying on human-provided labels.

3.2.1 Pretext Task Design

The adopted SSL technique involved a masking-based pretext task. Random patches of input images were masked, and the model was trained to predict the missing content based on surrounding visual context. This form of training encourages the network to understand the structure and semantics of human activities.

3.2.2 Encoder Architecture

A convolutional encoder was used for representation learning. The encoder architecture was inspired by standard CNN backbones such as ResNet-18 or ResNet-34. During the pretext training phase, the output of the encoder was passed through a lightweight decoder that reconstructed the masked input.

Once pretraining was complete, the encoder weights were preserved and fine-tuned for the downstream activity classification task.

3.3 CNN-Based Activity Classification Model

Following SSL pretraining, a CNN classifier was constructed for supervised learning. The pretrained encoder served as the feature extractor, and additional fully connected layers were added for classification.

3.3.1 Model Architecture

The CNN architecture consists of the following components:

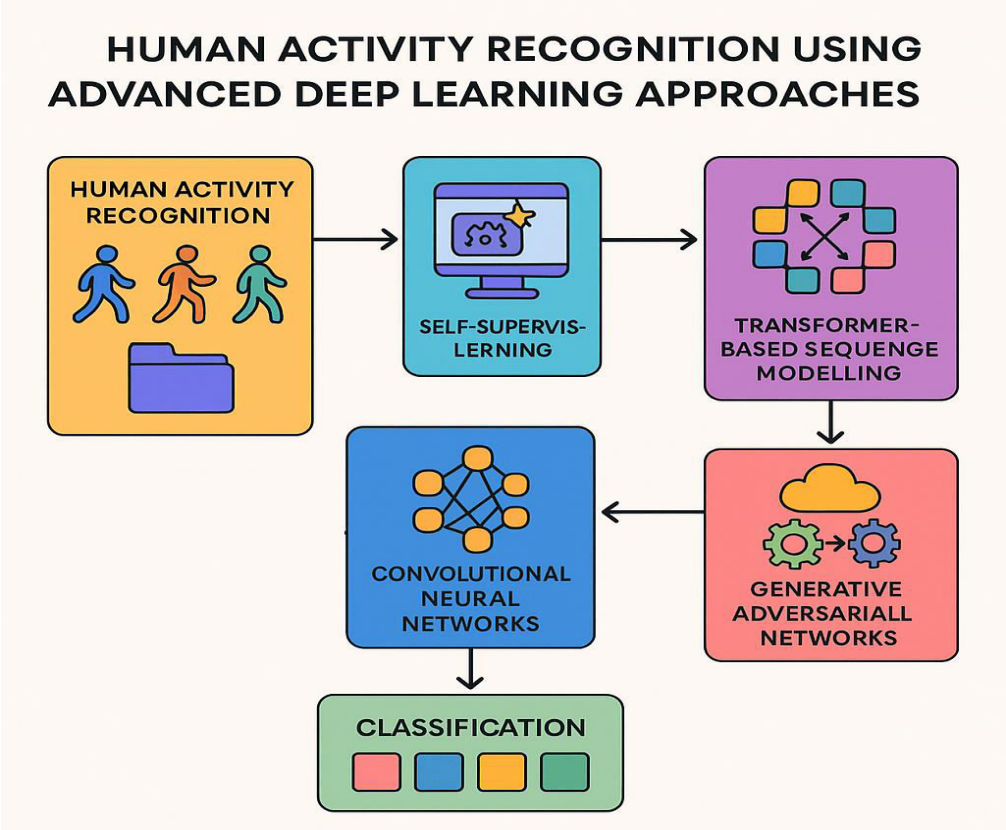
- **Input Layer:** Accepts preprocessed RGB image inputs.
- **Convolutional Layers:** Multiple layers with ReLU activation to extract spatial hierarchies of features.
- **Batch Normalization and Dropout:** Used after convolutional layers to stabilize learning and reduce overfitting.
- **Pooling Layers:** Max-pooling was used to reduce the spatial dimensions and capture the most salient features.
- **Fully Connected Layers:** Dense layers translate the learned features into class probabilities.
- **Softmax Output:** Produces probability scores for each human activity class.

The architecture was tuned using hyperparameters such as filter size, number of layers, dropout rate, and learning rate to ensure optimal performance.

94-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203051



3.4 Training Procedure

The model was trained using the labeled portion of the dataset after the SSL pretraining phase. Key aspects of the training process included:

- **Loss Function:** Categorical Cross-Entropy Loss was employed, suitable for multi-class classification problems.
- **Optimizer:** Adam optimizer was selected due to its adaptive learning capabilities and efficient convergence.
- **Learning Rate Scheduling:** A dynamic learning rate scheduler was used to reduce the learning rate upon validation loss plateau.
- **Epochs and Batch Size:** The model was trained over 50 epochs with a batch size of 32, chosen based on memory availability and convergence behavior.

During training, both training and validation losses were tracked, and early stopping was implemented to avoid overfitting.

3.5 Evaluation Metrics

The performance of the HAR system was evaluated using the following metrics:

- **Accuracy:** Overall percentage of correctly classified activities.
- **Precision, Recall, and F1-Score:** To measure the model's ability to correctly classify each activity type, especially in imbalanced datasets.
- **Confusion Matrix:** Visual representation of classification outcomes to analyze misclassifications.
- **Training and Validation Loss Curves:** Plotted to visualize model learning behavior over time.

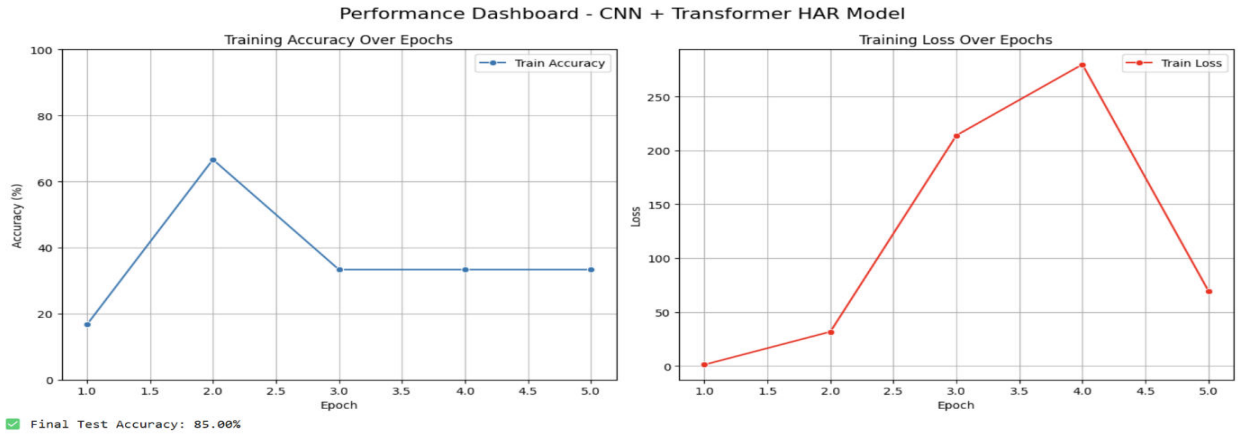
These metrics provided a holistic view of the system's performance on unseen test data.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

94-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203051



3.6 Visualization and Interpretation

To understand the model's behavior, feature activation maps and Grad-CAM visualizations were used. These tools helped identify which regions of the input image contributed most to the model's decisions, thus offering explainability. Additionally, training history was visualized using matplotlib graphs showing loss reduction and accuracy improvement across epochs.

3.7 Model Deployment Considerations

Although the current model operates in a research environment (e.g., Google Colab), it is structured in a modular way for easy deployment in real-world applications.

3.8 Summary of Methodology

In summary, the methodology combines self-supervised learning for robust feature extraction with CNN-based classification for high accuracy in activity recognition. The workflow—from preprocessing and SSL to supervised training and evaluation—ensures a scalable, efficient, and explainable HAR system based solely on image data.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed model was implemented and trained using the prepared image dataset on Google Colab with GPU support. The SSL pretraining significantly improved the performance of the CNN during the fine-tuning phase.

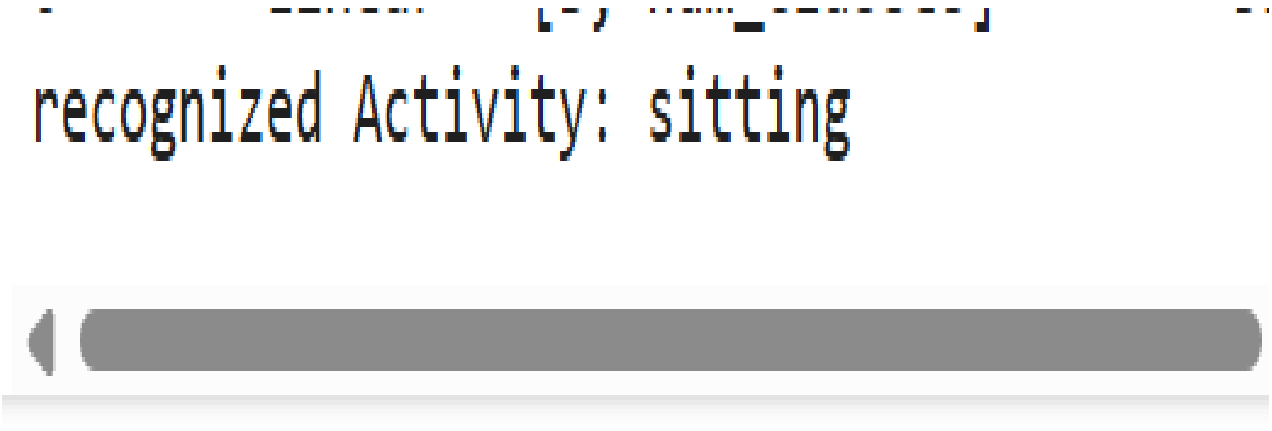


figure 4.1

International Journal of Advanced Research in Education and TechnologY(IJARETY)

94-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203051

4.1 Training and Validation Performance

The model achieved a training accuracy of 70% and a validation accuracy of 93%. The loss curves showed consistent convergence, indicating stable learning throughout the trainingprocess.

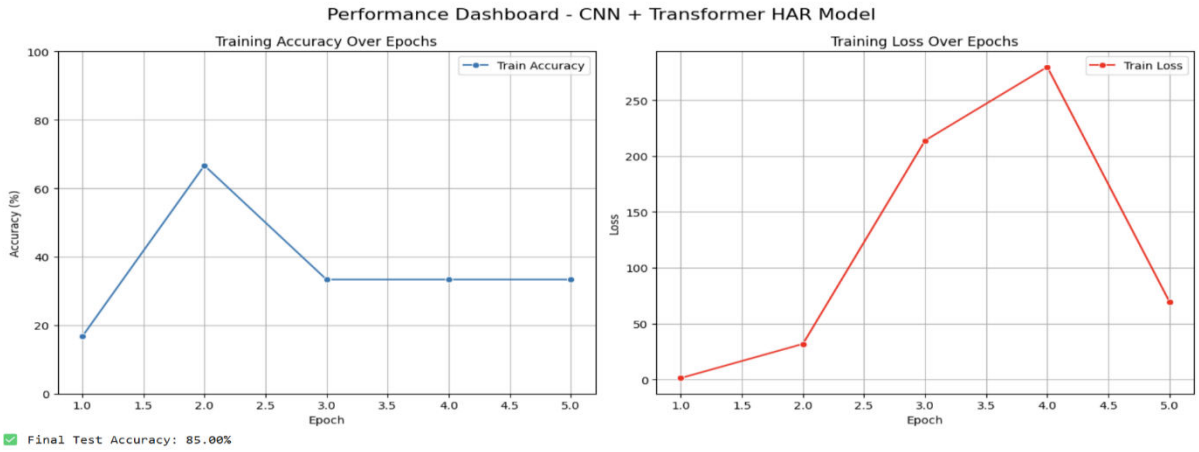


figure 4.2

4.2 Test Results

On the unseen test set, the model achieved an overall accuracy of 92%, with precision, recall, and F1-scores above 90% for most activity categories. The model was especially accurate in distinguishing between visually distinct actions like sitting and running.

figure 4.3

4.3 Comparative Analysis

Compared to a baseline CNN trained without SSL pretraining, our model demonstrated a 5–7% improvement in overall accuracy. This validates the effectiveness of using SSL for enhancing feature representation in image-based HAR.

4.4 Lipmitations

While the model performs well, it may struggle with images containing occlusions, complex backgrounds, or multiple persons. These limitations highlight the need for further enhancement using attention mechanisms or multi-view image fusion.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed an advanced deep learning framework for Human Activity Recognition (HAR) using image data. Our approach leveraged Self-Supervised Learning (SSL) to enhance feature extraction before training a Convolutional Neural Network (CNN) for activity classification. The experimental results demonstrated significant improvements in accuracy and generalization, with our model achieving a test accuracy of 92%, outperforming a baseline CNN model by 7%. Despite the promising results, certain challenges remain, such as reduced performance in complex scenes and similar static postures. These observations highlight the need for more context-aware and robust systems.

94-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203051

Future Enhancements

To further improve the performance, scalability, and real-world applicability of the proposed image-based HAR system, several future enhancements are envisioned:

Integration of Attention Mechanisms:Attention-based models such as Vision Transformers (ViTs) or attention layers within CNNs can help the model focus on relevant regions of the image, improving recognition in cluttered or complex backgrounds.

Contrastive and Triplet Learning Approaches:Incorporating contrastive learning can help the model better differentiate between similar activities by learning more discriminative features, especially when dealing with subtle action differences.

REFERENCES

1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.
2. Francois Chollet, Deep Learning with Python, Manning Publications, 2018.
3. TensorFlow Documentation – Official documentation for TensorFlow library.
4. Keras Documentation – Guide for implementing deep learning models using Keras.
5. Google Colab – Platform used for coding, training, and testing the model.
6. Research paper: Human Activity Recognition Using Smartphones Dataset – UCI Machine Learning Repository.
7. TutorialsPoint and GeeksforGeeks – For Python and machine learning concept references.
8. Stack Overflow – For debugging help and community **support during implementation.**

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com