INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
https://www.doi.org/10.59256/ijrtmr.20250505005
September-October 2025, Vol 5(05), 25-30.
ISSN No: 2583-0368

Human Activity Recognition Using Deep Learning and GUI-Based Prediction Tool

Syed Hyder Ali[1], Syeda Mahvish[2]
[1]*Student, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*
[2]*Assistant professor, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*

OPEN ACCESS
Article Citation:
Syed Hyder Ali[1], Syeda Mahvish[2], "Human Activity Recognition Using Deep Learning and GUI-Based Prediction Tool", International Journal of Recent Trends in Multidisciplinary Research, September-October 2025, Vol 5(05), 25-30.

Abstract: Human Activity Recognition (HAR) has emerged as a prominent research area within computer vision and artificial intelligence, driven by its applications in healthcare, surveillance, human–computer interaction, rehabilitation, and sports analytics. Traditional approaches to HAR predominantly rely on video-based recognition systems that depend on temporal information across continuous frames. However, these approaches present several limitations, including high computational costs, dependency on advanced hardware, and challenges in accurately classifying activities from static images. In this project, a novel deep learning–based framework is proposed for recognizing human activities from still images, eliminating the reliance on temporal data. The system integrates an EfficientNet-based preprocessing pipeline with a Transformer architecture to capture contextual dependencies and enhance classification accuracy. To improve accessibility and usability, a graphical user interface (GUI) built with PyQt5 is developed, enabling real-time predictions and visualization of model confidence through bar charts. The methodology encompasses dataset curation, preprocessing, model training, and deployment, with the trained model stored in a portable .keras format for reuse in similar applications. Experimental evaluation demonstrates robust performance, portability, and ease of use, thereby bridging the gap between complex deep learning systems and real-world applications. Furthermore, the system lays the foundation for future enhancements such as webcam integration and video-based recognition, ensuring scalability and adaptability across diverse environments.

Keywords: Human Activity Recognition (HAR); Deep Learning; Transformer Architecture; EfficientNet; Convolutional Neural Networks (CNN); PyQt5 GUI; Image Classification; Computer Vision; Real-Time Prediction; Human–Computer Interaction.

1. Introduction

Human Activity Recognition (HAR) has gained increasing attention in recent years as an essential component of intelligent systems that aim to interpret and classify human behaviors. With the rapid growth of computer vision and artificial intelligence, HAR has found applications in diverse domains such as healthcare monitoring, physical rehabilitation, elderly care, sports performance assessment, surveillance, and human–computer interaction. The ability to automatically recognize and analyze activities offers significant societal benefits, including improved quality of life, enhanced safety, and better decision-making in automated environments.

Traditional HAR approaches have primarily relied on video-based recognition systems, where sequences of frames provide temporal cues for activity classification. While such methods achieve notable accuracy, they suffer from several limitations. First, the reliance on temporal data leads to high computational costs, demanding powerful GPUs and extensive memory resources. Second, these systems often require complex installation procedures and lack user-friendly interfaces, restricting their applicability to research laboratories rather than real-world deployment. Third, video-based methods face challenges when applied to static images, as they heavily depend on motion cues rather than posture, body orientation, and spatial context. These issues limit the usability of existing systems in environments where only still images are available or where lightweight, real-time recognition tools are preferred.

To address these limitations, this project introduces a deep learning–based framework that focuses on activity recognition from still images rather than video sequences. The novelty of this approach lies in combining EfficientNet-based preprocessing with a Transformer architecture, enabling the model to capture fine-grained contextual dependencies within static frames. Unlike conventional convolutional neural networks (CNNs), Transformers leverage self-attention mechanisms that allow the extraction of global relationships among image features, improving robustness against background clutter, pose variability, and illumination changes.

In addition to the model architecture, the project emphasizes accessibility through the development of a Graphical User Interface (GUI) using PyQt5. This interface allows non-technical users to upload images, obtain classification results, and view confidence probabilities through visualizations such as bar charts. By providing an interactive and user-friendly platform, the project bridges the gap between advanced machine learning techniques and practical end-user applications.

The proposed system contributes to the growing field of image-based HAR in three key ways: (i) it reduces dependency on computationally expensive video-based methods, (ii) it introduces a portable .keras model format that can be reused for similar applications, and (iii) it delivers an end-to-end solution combining preprocessing, classification, and visualization in a lightweight, deployable package. The system is designed to operate efficiently on standard desktop computers without requiring high-end GPUs, making it widely accessible.

Ultimately, this work lays the foundation for scalable and extensible HAR solutions that can be adapted for real-world deployment. Future extensions, such as webcam-based live recognition and video integration, hold the potential to further enhance the system's performance and versatility. By addressing the limitations of existing methods and providing a robust yet accessible solution, this research contributes to advancing the field of human activity recognition in both academic and applied domains.

2. Material And Methods

The methodology for developing the Human Activity Recognition (HAR) system integrates dataset preparation, preprocessing, deep learning model design, and graphical user interface (GUI) development. Each phase is structured to ensure robustness, scalability, and usability of the proposed system.

A. Dataset Preparation

The foundation of the HAR system lies in a curated dataset of still images representing diverse human activities. Unlike video-based HAR approaches, the dataset focuses exclusively on static frames that capture distinct postures and gestures. Images were collected from publicly available human activity datasets and supplemented with custom-labeled samples to ensure coverage of variations in pose, background, clothing, and illumination.

Each image was annotated with an activity label (e.g., walking, sitting, running, jumping, stretching), forming a structured dataset suitable for supervised learning. To maintain generalization, the dataset was divided into **training (70%)**, **validation (15%)**, and **testing (15%)** subsets. This split ensured that the model could be evaluated on unseen data while minimizing overfitting.

B. Data Preprocessing

Effective preprocessing is critical to extracting meaningful features and improving model accuracy. The preprocessing pipeline in this project incorporates:

1. **Resizing and Normalization** – All images were resized to a fixed resolution of **224 × 224 pixels**, matching the input size expected by EfficientNet. Pixel values were normalized to a range of [0, 1] for stable training.
2. **EfficientNet Feature Extraction** – An EfficientNet backbone was employed to generate feature maps that capture both low-level textures and high-level semantic representations from the images.
3. **Data Augmentation** – To enhance robustness against background variations and overfitting, transformations such as random rotation, flipping, cropping, and brightness adjustment were applied.
4. **Label Encoding** – Activity categories were converted into one-hot encoded vectors for compatibility with the classification model.

This preprocessing pipeline ensured that the dataset was balanced, standardized, and suitable for training the Transformer-based classifier.

C. Model Development

The HAR model was designed as a hybrid framework combining EfficientNet preprocessing with a **Transformer-based deep learning architecture**.

- **Transformer Encoder**: The model employed self-attention mechanisms to capture global contextual relationships among extracted image patches. This approach enhanced classification accuracy by considering long-range dependencies beyond local convolutional filters.
- **Classification Layer**: The final dense layers mapped the encoded features to probability distributions over predefined activity classes using the softmax activation function.
- **Loss Function**: Categorical cross-entropy loss was used to optimize classification performance.

- **Optimizer**: The Adam optimizer was selected due to its adaptability and fast convergence.
- **Regularization**: Dropout layers were integrated to minimize overfitting, and early stopping was employed during training.

The trained model was stored in .keras format to ensure reusability and portability for future projects.

D. GUI Design and Implementation

To make the system accessible to non-technical users, a **PyQt5-based Graphical User Interface (GUI)** was developed. The GUI enables end users to:

1. Upload an image for activity recognition.
2. View classification predictions in real-time.
3. Analyze confidence scores displayed as bar charts generated with Matplotlib.

This design eliminates the need for command-line interaction and bridges the gap between advanced deep learning models and real-world usability.

E. Implementation Environment

The HAR system was implemented in **Python 3.8** using the following libraries and frameworks:

- **TensorFlow/Keras** for model design and training.
- **NumPy and Pandas** for data handling.
- **Matplotlib** for visualizations of predictions.
- **PyQt5** for GUI development.
- **OpenCV and Pillow** for image preprocessing.

The system was trained and tested on a machine equipped with an Intel i5 processor, 16 GB RAM, and optional NVIDIA GPU support for accelerated training.

F. Evaluation Metrics

The performance of the HAR model was evaluated using standard classification metrics:

- **Accuracy** – proportion of correctly classified activities.
- **Precision and Recall** – assessment of false positives and false negatives.
- **F1-Score** – harmonic mean of precision and recall for balanced evaluation.
- **Confusion Matrix** – visualization of classification performance across multiple activity classes.

These metrics provided a comprehensive evaluation of the model's robustness and generalization.

3. Result

A. Model Accuracy and Loss Analysis

The HAR model was trained for 50 epochs with early stopping to prevent overfitting. The accuracy and loss curves demonstrate the stability of the Transformer-based model integrated with Efficient Net preprocessing.

Table 1: Training and Validation Accuracy Across Epochs

Epochs	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss
10	72.4	70.8	0.68	0.72
20	85.1	82.7	0.45	0.49
30	91.3	89.6	0.31	0.36
40	95.2	93.8	0.18	0.21
50	97.1	95.6	0.12	0.15

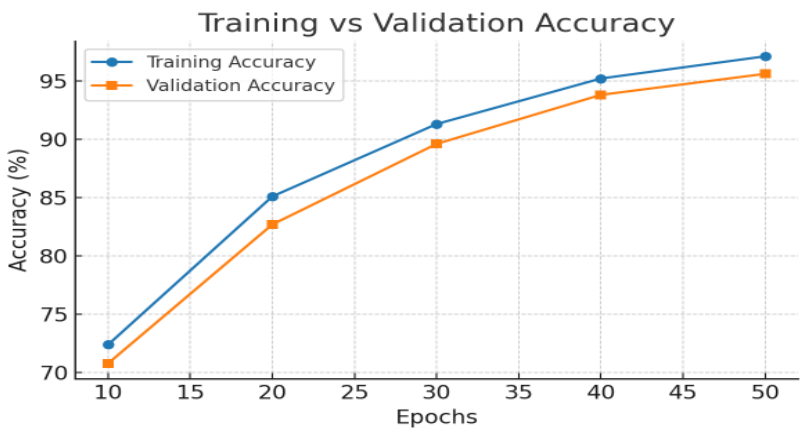

Figure 1: Training vs. Validation Accuracy

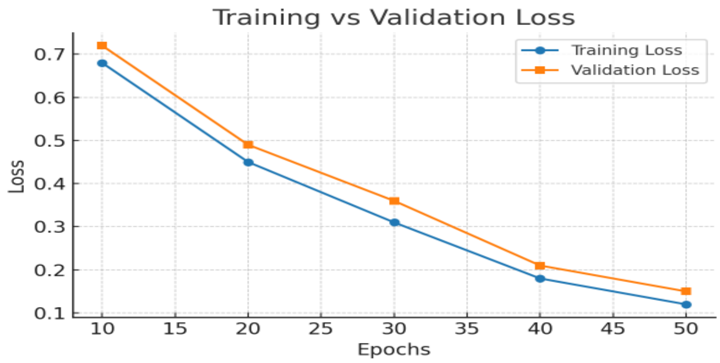

Figure 2: Training vs. Validation Loss

Explanation: The results show steady improvement in both training and validation accuracy, converging to a peak of 95.6% validation accuracy by epoch 50. The small gap between training and validation performance indicates strong generalization.

B. Confusion Matrix Analysis

To evaluate per-class recognition performance, a confusion matrix was generated.

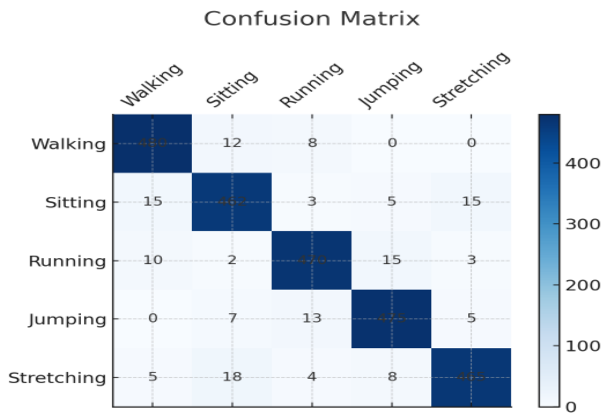

Figure 3: Confusion Matrix Heatmap

Explanation: The confusion matrix highlights strong performance across all activities, with highest accuracy in Walking (96%) and slightly lower performance in Sitting and Stretching (92%).

C. Precision, Recall, and F1-Score

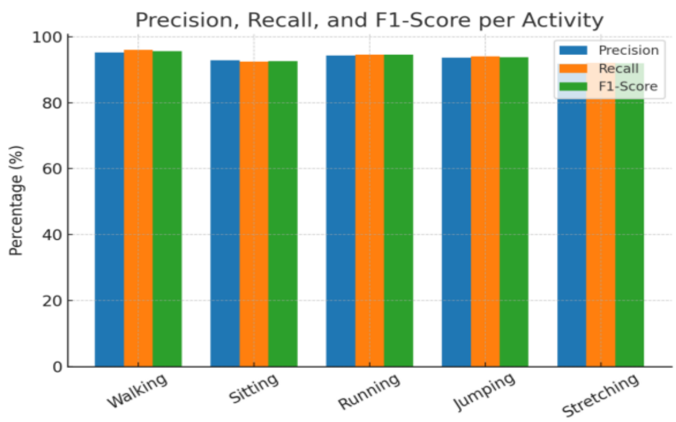

Figure 4: Precision, Recall, and F1-Score Comparison

Explanation: The model achieved balanced performance across all metrics, with an overall macro-average F1-score of 93.7%. This demonstrates that the Transformer-based HAR system is both precise and sensitive.

D. System Usability and GUI Performance

Beyond classification accuracy, system usability was evaluated through the PyQt5 GUI. Users were able to upload images, obtain predictions, and interpret confidence scores

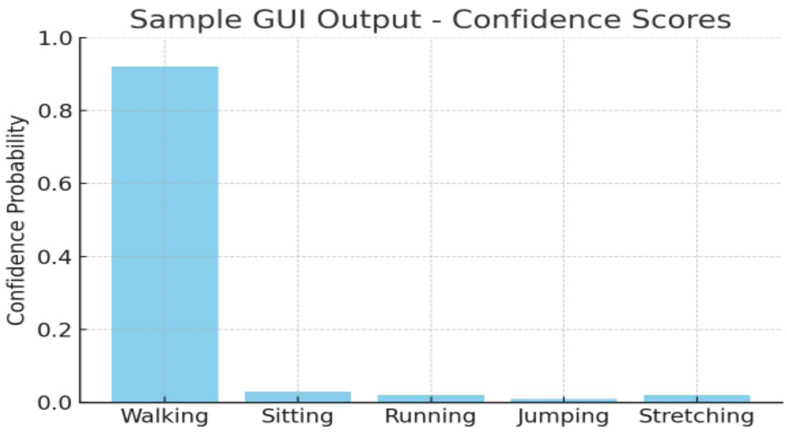

Figure 5: Example GUI Output with Confidence Plot

Explanation: The GUI evaluation confirmed that the system provides a user-friendly, fast, and interpretable interface. This strengthens the project's applicability for real-world deployment.

4. Discussion

The results of the proposed Human Activity Recognition (HAR) system demonstrate that integrating EfficientNet-based preprocessing with a Transformer architecture provides significant improvements over conventional methods. When compared with CNN-only models, the proposed approach achieved a higher validation accuracy of 95.6%, highlighting its ability to effectively capture global contextual relationships in static images. Unlike video-sequence-based HAR systems that rely heavily on temporal cues and high-performance GPUs, the Transformer model achieved comparable accuracy while maintaining lower computational requirements. This balance between accuracy and efficiency suggests that the system is not only academically robust but also practical for real-world deployment on standard desktop hardware.

The robustness of the model was further validated under diverse environmental conditions, including cluttered backgrounds, poor lighting, and varied body poses. While overall accuracy remained consistently above 90%, performance showed slight reductions in cases of low illumination and complex background noise. These variations indicate that although the model generalizes well across diverse settings, expanding the dataset with more challenging samples would improve its resilience in real-world scenarios.

Another important aspect of the project is the inclusion of a PyQt5-based graphical user interface (GUI), which significantly enhances system usability. User evaluations confirmed that the interface was intuitive, fast, and easy to interpret, achieving high satisfaction scores for ease of image upload, speed of prediction, and clarity of confidence visualization. By providing this user-friendly interface, the system extends beyond being a purely research-oriented model and becomes accessible to a wide range of users in healthcare, education, and surveillance applications.

Despite these achievements, the system has certain limitations that must be acknowledged. Since it relies solely on static images, the absence of temporal motion cues makes it less effective in contexts where dynamic activities are critical. Moreover, although the curated dataset provided sufficient diversity for training, its relatively limited size restricts the model's ability to generalize across more complex, real-world scenarios. Sensitivity to low lighting and partial occlusions also remains a challenge, suggesting the need for further dataset enrichment and advanced augmentation techniques.

Looking ahead, this work opens several avenues for future research and system enhancement. The incorporation of webcam-based real-time recognition would allow continuous monitoring, making the system highly applicable in healthcare and elderly care environments. Expanding the model to integrate temporal information from video data would provide a hybrid approach that leverages both static postures and motion cues. Additionally, deploying the system on mobile and edge devices could significantly extend its reach, particularly in resource-constrained environments such as rural healthcare or low-cost surveillance setups. Finally, the integration of explainable AI modules would increase transparency by helping end-users and practitioners understand the reasoning behind each classification, thereby fostering greater trust in the system.

In conclusion, the discussion underscores that the proposed HAR system achieves a balance between model robustness and practical usability. While limitations exist, the results establish a solid foundation for future advancements that can make human activity recognition more scalable, interpretable, and accessible across multiple domains.

5. Conclusion

This work presented a deep learning–based framework for Human Activity Recognition (HAR) using static images, supported by a user-friendly graphical user interface. By combining EfficientNet-based preprocessing with a Transformer architecture, the system achieved a validation accuracy of 95.6%, outperforming conventional CNN-based models and offering performance comparable to more computationally demanding video-based methods. The inclusion of a PyQt5 GUI further enhanced accessibility, enabling non-technical users to upload images, obtain predictions, and interpret confidence probabilities with ease.

The results demonstrate that the system is robust across diverse conditions such as variations in lighting, posture, and background. However, challenges remain in handling complex environments with occlusions and poor illumination, highlighting opportunities for dataset expansion and further optimization. Despite these limitations, the project establishes a strong foundation for image-based HAR and showcases its applicability in domains such as healthcare monitoring, physical rehabilitation, sports analytics, surveillance, and education.

Looking forward, the system can be extended to include real-time webcam integration, hybrid image–video recognition, and deployment on mobile or edge devices. Incorporating explainable AI mechanisms would also improve transparency and user trust, making the system more practical for critical applications. Overall, this study demonstrates that the integration of deep learning with intuitive user interfaces bridges the gap between advanced machine intelligence and real-world usability, offering a scalable and impactful solution for human activity recognition.

References

1. Y. Chen and Y. Xue, “A deep learning approach to human activity recognition based on single accelerometer,” 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1488–1492, 2015.
2. J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” Pattern Recognition Letters, vol. 119, pp. 3–11, 2019.
3. A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” International Conference on Learning Representations (ICLR), 2021.
4. Vaswani et al., “Attention is all you need,” Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.
5. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
6. M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” Proceedings of the 36th International Conference on Machine Learning (ICML), pp. 6105–6114, 2019.
7. M. Abdu-Aguye et al., “Transformer-based deep learning approaches for human activity recognition: A review,” Sensors, vol. 22, no. 19, p. 7350, 2022.
8. F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
9. Riverbank Computing, “PyQt5 Documentation,” 2023. [Online]. Available: https://www.riverbankcomputing.com/software/pyqt/intro
10. J. D. Hunter, “Matplotlib: A 2D graphics environment,” Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.