International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

A Deep Learning Approach for Human Activity Recognition in Healthcare Applications

Mrs K. Anusha[1]|Aalla Hema Sri Lakshmi[2]|Putta Gopala Swami[3]|Abburi Srinivasa Rao[4]|Bellamkonda Sai Sandeep[5].

[1,2,3,4,5]Department of CSE-AIML, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India

ABSTRACT: Human Activity Recognition (HAR) has become an important area of research due to its wide applications in surveillance, healthcare, and human computer interaction, where accurate understanding of human behavior is essential. However, traditional approaches and basic convolutional neural networks often face limitations in capturing complex visual patterns and require large amounts of data for effective performance. This project aims to develop an efficient and accurate HAR system by leveraging transfer learning with the EfficientNetB0 model and comparing its performance with a conventional CNN approach. The proposed methodology utilizes EfficientNetB0 as a pre-trained feature extractor to learn high-level spatial features from input images or video frames, followed by fully connected layers for classification, while the existing CNN model is trained from scratch for performance comparison. Experimental results demonstrate that the proposed EfficientNetB0-based model achieves higher accuracy, compared to the CNN model, while also reducing training time and improving generalization. The study concludes that integrating transfer learning with EfficientNetB0 significantly enhances the effectiveness of human activity recognition systems, making it a reliable and scalable solution for real-world applications.

Keywords: Human Activity Recognition, Convolutional Neural Network, EfficientNetB0, Deep Learning, Computer Vision, Activity Classification.

1. INTRODUCTION

Human Activity Recognition (HAR) has emerged as a critical research domain due to its wide applicability in surveillance systems, healthcare monitoring, smart environments, and human–computer interaction. The ability to automatically identify human actions from visual data enables intelligent decision-making and real-time behavioral analysis. Earlier approaches relied on handcrafted features combined with traditional machine learning techniques, which struggled to represent complex visual patterns and variations in human activities. The advancement of deep learning, particularly Convolutional Neural Networks, has significantly improved feature extraction by learning hierarchical representations directly from data. However, training deep networks from scratch requires large-scale datasets and high computational resources. Transfer learning has addressed these challenges by enabling models pre-trained on large datasets to be adapted for specific tasks with improved efficiency and performance. EfficientNetB0, a modern deep learning architecture, provides a balanced scaling of network depth, width, and resolution, leading to better accuracy with fewer parameters. By leveraging transfer learning with EfficientNetB0, human activity recognition systems can achieve higher precision and faster convergence. The integration of such optimized models enhances the reliability of activity classification in real-world environments.

2. LITERATURE REVIEW

Arun Kumar Jhapate [2] targets to make a system that can detect unusual activities in the crowd and that too at real-time. The concept is fairly new in the field of research as very less work is done on it. It proves to be promising as it provides an accuracy of 96.42%.

Cha Zhang and Zhengyou Zhang e [3] proposed a method that can detect faces at a fast pace and can also learn at the same time. The Deep Convolutional Neural Network (DCNN) uses a Divide & conquer approach with the help of which the faces can be divided into multiple categories which help in making decisions as if a face is present in an image, its pose, etc.

Adrian Rosebrock [4] in his work discusses the whole process of building a face detection model using deep learning with the help of Python and OpenCV. The model uses a special technique known as deep metric learning. It uses libraries like dlib and face recognition. The results are both accurate and capable of being executed on GPU in real-time as well.

Yang Weilong [5] proposed an approach that uses still images for recognizing human activities. The proposed system combines the pose estimation with action recognition which. can further be used in video input too. While training, the author makes 90 poses for every pose, the SVM classifier is trained based on the HOG descriptors.

International Journal of Engineering Research and Science & Technology

Eshaan Ramnani, and Sanjay Singh [6] in their paper elaborate on the extended scope of Multi view Face Detection using CNN. The face detection activity is performed by the Deep Dense Face Detection, which further works on a single model that relies on DCNN. Some related work for face detection and face recognition has been done using facial landmarks. The application of human activity recognition can be seen in human interpersonal relations and their interaction. Some of the human's natural activities like eating and running can be recognized with ease. The complex task is normally sub-divided into smaller tasks that are easy to recognize. Knowledge of the surroundings of the object in real-time can help the system to know and understand the Human Activity with ease and time efficiently.

3. SYSTEM ANALYSIS

Human action recognition based on deep learning typically involves a multi-stage system architecture designed to process video data and classify human actions. The system begins by capturing video frames or sequences, which are then fed into a Convolutional Neural Network (CNN) for feature extraction, detecting key elements like body parts, motion, and objects. For temporal analysis, EfficientnetB0 The deep learning model is trained on labeled action datasets, and once trained, it can accurately recognize and classify human actions in real-time or from video recordings. This approach leverages the power of deep learning to capture both spatial and temporal dependencies, making it highly effective for applications like surveillance, human-computer interaction, and video analytics

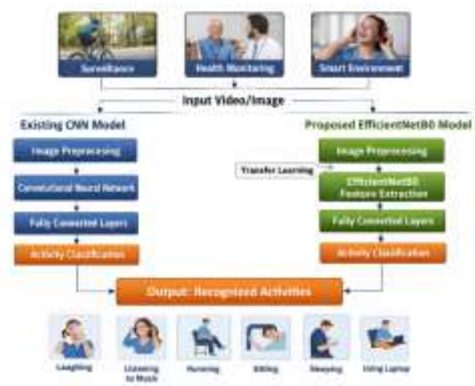

Fig 1: system architecture

3.1 Methodology

The methodology of the proposed Human Activity Recognition system follows a structured pipeline consisting of dataset acquisition, preprocessing, model development, evaluation, and real-time detection using both existing and proposed algorithms. Initially, a labeled dataset containing multiple human activity classes is collected from standard image and video sources, ensuring diversity in pose, background, and lighting conditions. In the preprocessing stage, raw data is cleaned, resized to a fixed dimension, normalized, and augmented using techniques such as rotation, flipping, and scaling to improve generalization. For model development, a baseline Convolutional Neural Network (CNN) is implemented to extract spatial features, while the proposed EfficientNetB0-based transfer learning model is developed by fine-tuning pre-trained weights to capture more complex and high-level representations. Both models are trained using the prepared dataset with optimized hyperparameters and loss functions. In the evaluation stage, performance is measured using accuracy and loss metrics, along with validation techniques to compare the effectiveness of the existing and proposed models, where the EfficientNetB0 model demonstrates superior accuracy and faster convergence. Finally, the trained model is integrated into a real-time detection system that processes live input images or video frames to recognize and classify human activities dynamically, ensuring practical applicability in real-world scenarios.

3.2 **Dataset Collection**: This module gathers and organizes labelled human activity images for training and testing. The dataset is cleaned, resized, and split into training and validation sets using stratified sampling to maintain balanced class distribution.

3.3**Data Pre-processing**: This module prepares images for model input by resizing them to 224×224 pixels and normalizing pixel values. Data augmentation techniques like flipping, rotation, and zooming are applied to improve generalization and robustness.

3.3**Feature Extraction Module** :This module uses EfficientNetB0 as a pre-trained transfer learning model to extract high-level spatial features from images. Global average pooling and fine-tuning are applied to adapt the model for human activity recognition.

3.4 **Model Building**: This module builds a baseline CNN model trained from scratch to learn spatial features directly from the dataset. It includes convolutional, pooling, and dense layers and is used for performance comparison with EfficientNetB0

3.5 Prediction activity : This module performs final prediction using the trained EfficientNetB0 model. A softmax output layer assigns probabilities to each activity class, and the highest probability determines the predicted human activity.

4. Design and Construction

The design and construction of the proposed Human Activity Recognition system are centered around an EfficientNetB0-based deep learning architecture integrated into a scalable and efficient processing pipeline.

The system is designed with multiple modules including data input, preprocessing, feature extraction, classification, and real-time detection. Initially, input images or video frames are collected and passed through a preprocessing module where resizing, normalization, and augmentation are applied to standardize the data. The core component of the system is the EfficientNetB0 model, which acts as a feature extractor using pre-trained weights and fine-tuned layers to capture high-level spatial features such as human posture and motion context. The extracted features are then passed through global pooling, batch normalization, and fully connected layers, followed by a soft max classifier to predict activity classes. The construction of the model is implemented using deep learning frameworks such as Tensor Flow or Keras, ensuring modularity and flexibility for training and deployment. Additionally, the system incorporates optimization techniques, dropout regularization, and early stopping mechanisms to enhance performance and prevent overfitting. Finally, the trained model is deployed into a real-time environment where it processes live input streams and accurately recognizes human activities, making the system robust, efficient, and suitable for practical applications.

5. RESULTS AND DISCUSSION

The results of the proposed human activity recognition system demonstrate that the use of EfficientNetB0 with transfer learning provides effective and accurate classification of human activities from image data. During training, the model shows a consistent decrease in loss values along with a steady increase in accuracy, indicating successful learning of meaningful spatial features. The training and validation performance curves confirm that the model improves across epochs and maintains stable performance on validation data, reflecting strong generalization capability. The EfficientNetB0 model extracts rich and high-level visual features such as posture, object interaction, and contextual information, which significantly enhances classification performance. In comparison, the conventional CNN model achieves lower accuracy and requires more training time, highlighting the efficiency of the proposed approach. The best model checkpoint, selected based on validation accuracy, ensures reliable prediction across different activity classes.

Fig2: Environment

Figure 2 shows Command Prompt to Run the Application. The application runs on a local server (http://127.0.0.1:5000)

Fig 3: Register page

Fig 3 shows the user Register page where users enter their email and password to access the system securely. It ensures authentication and provides access only to authorized users.

Fig 4: prediction

Figure 4 shows the "Predict Image" page of a Human Activity Recognition (HAR) system. A file explorer window is open, allowing the user to select an image from a dataset. The system is prepared to upload the chosen image and predict the activity using the model.

Fig 5: Predicted Output

Fig 5 shows a Human Activity Recognition (HAR) system interface used for predicting actions from images .A user uploads an image, and the system analyzes it using a model to detect the activity. the system predicts the activity as "sleeping" based on the uploaded image.

Fig 6: EDA and metrics

Fig 6 shows a dashboard of a Human Activity Recognition (HAR) system displaying model performance. It includes graphs for accuracy and loss over training epochs, along with class distribution of activities. The model used is EfficientNetB0, achieving an accuracy of about 97.5%.

6. CONCLUSION

The developed human activity recognition system demonstrates that the integration of transfer learning with EfficientNetB0 significantly improves the accuracy and efficiency of activity classification from image data. The model effectively learns high-level spatial features and achieves strong performance compared to a conventional CNN approach. The training process shows stable convergence with reduced loss and increased accuracy, indicating reliable learning behavior. The system also generalizes well to unseen data, making it suitable for real-world applications such as surveillance, healthcare monitoring, and smart environments. The comparison with the existing CNN model confirms that the proposed approach reduces training complexity while delivering higher accuracy and better performance.

FUTURE SCOPE: Future work of the Human Activity Recognition system can focus on integrating multimodal data such as wearable sensors, depth information, and audio signals to improve recognition accuracy. Advanced deep learning models like LSTM and Transformers can enhance temporal motion analysis for complex activities. will enable real-time, scalable, and adaptive activity monitoring in real-world applications.

REFERENCES

[1] Vrigkas, C. Nikou, and I. A. Kakadiaris, “A review of human activity recognition methods,” Frontiers in Robotics and AI, vol. 2, p. 28, 2015.

[2] Jhapate, Arun Kumar, Sunil Malviya, and Monika Jhapate., “Unusual Crowd Activity Detection using OpenCV and Motion Influence Map,” 2nd International Conference on Data, Engineering and Applications (IDEA), pp. 1-6, IEEE, 2020.

[3] Zhang, Cha, and Zhengyou Zhang, “Improving multiview face detection with multi-task deep convolutional neural networks,” In IEEE Winter Conference on Applications of Computer Vision, pp. 1036-1041, IEEE, 2014.

[4] Triantafyllidou, Danai, and Anastasios Tefas. “Face detection based on deep convolutional neural networks exploiting incremental facial part learning,” In 23rd International Conference on Pattern Recognition (ICPR), pp. 3560-3565, IEEE, 2016.

[5] Yang, Weilong, Yang Wang, and Greg Mori, “Recognizing human actions from still images with latent poses,” In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2030-2037, IEEE, 2010.

[6] Mehta, Jinesh, Eshaan Ramnani, and Sanjay Singh, “Face detection and tagging using deep learning,” In 2018 International Conference on Computer, Communication, and Signal Processing (ICCCSP), pp. 1-6, IEEE, 2018.