International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(4): 174-186 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i04.01393

Machine Learning Method for Understanding Human Behaviour and Human Action Recognition

Anand Deva Durai C*, Abeer Almansour

Department of Computer Science, College of Computer Science, King Khalid University, Abha, Saudi Arabia. *Corresponding Author's Email: anandevadurai@kku.edu.sa

Abstract

Human behaviour and action recognition are vital components of effective surveillance video analysis, playing a key role in maintaining public safety. Current approaches, such as 3D Convolutional Neural Networks (3D CNN) and two-stream neural networks (2SNN), often struggle with computational inefficiencies due to their high parameter demands. In response to these challenges, we introduce HARNet, a lightweight residual 3D CNN architecture based on directed acyclic graphs, specifically designed to enhance the efficiency of human action detection. HARNet employs a novel pipeline to generate spatial motion data from raw video inputs, enabling robust latent representation learning of human motion. Unlike traditional methods, our approach processes both spatial and motion information within a single stream, effectively utilizing both types of cues. To further improve the discriminative capability of the extracted features, we integrate a Support Vector Machine (SVM) classifier on the latent representations obtained from HARNet's fully connected layer. Comprehensive evaluations on the UCF101, HMDB51, and KTH datasets show significant performance gains of 2.75%, 10.94%, and 0.18%, respectively. These results highlight the strength of HARNet's streamlined design and the effectiveness of combining SVM classifiers with deep feature learning for accurate and efficient human action recognition in surveillance videos. This work advances the field of reliable video analysis for real-world applications.

Keywords: 3D Convolutional Neural Networks (3D CNN), Directed Acyclic Graphs, Human Action Recognition Network (HAR Net), Spatial Motion, Support Vector Machine (SVM).

Introduction

Human action recognition in surveillance videos is a critical task with broad applications across industries such as safety and security, healthcare, and human-computer interaction (1). The ability to understand and analyze human behavior through video footage is essential for enhancing situational awareness and ensuring safety. Actions like walking, running, or engaging in specific activities can provide valuable insights for identifying potential threats, detecting unusual behavior, and monitoring critical scenarios. Traditional approaches to human action recognition often relied on handcrafted features and shallow classifiers, which struggled to capture the complex and diverse patterns inherent in human actions (2). However, the advent of deep learning has revolutionized the field, leading to significant advancements in action recognition with the use of deep neural networks, particularly 3D Convolutional Neural Networks (3D CNNs) and two-stream networks (3). These models have excelled by leveraging both spatial and temporal information in video sequences, achieving impressive performance by incorporating motion cues. Despite their success, these methods come with computational challenges due to their large parameter sizes and the high computational cost required for training and deployment (4). This makes them less suitable for real-time applications or environments with limited resources. To address these challenges, we propose a novel approach named HARNet (Human Action Recognition Network), designed to enhance the efficiency of human action recognition while maintaining high accuracy. HARNet is a lightweight residual 3D CNN architecture built on directed acyclic graphs, specifically aimed at reducing computational overhead. The key innovation of our method lies in the development of a new pipeline that generates spatial motion data from raw video inputs (5). This pipeline enables the extraction of essential features that capture both spatial and motion information, facilitating effective latent

(Received 22nd June 2024; Accepted 21st October 2024; Published 30th October 2024)

representation learning of human actions. Furthermore, to enhance the discriminative power of the learned features, we integrate traditional machine learning classifiers into our framework. Specifically, we apply a Support Vector Machine (SVM) classifier to the deep features extracted from HARNet's fully connected layer, enabling more accurate action recognition (6). We conduct extensive experiments and compare HARNet's performance against state-of-the-art algorithms on well-established action recognition datasets, including UCF101, HMDB51, and KTH (7). Our results demonstrate that the proposed HARNet-SVM approach outperforms existing methods, showing notable improvements in recognition accuracy. This highlights the potential of our lightweight architecture to drive advancements in human action recognition, particularly in resource-constrained environments where computational efficiency is critical. Recent years have witnessed a surge in the use of deep learning networks for video analysis tasks (8). For instance, one approach used ResNet50 to extract frame-level features, followed by Conv STM for detecting anomalous events. Another method combined CNN and LSTM architectures to classify videos using smaller datasets, analyzing RGB frames, optical flow, and their fusion through ResNet-152 and a three-layer LSTM. In addition, the Hybrid Deep Learning Network (HDLN) was introduced to extract features from complex inertial data in smartphone-based activity recognition (9). Deep learning models have also been utilized for worker activity recognition, integrating CNN, SVM, and R-CNN to extract fine-grained action features using an action-independent Gaussian mixture model (AIGMM). Other research has focused on pedestrian attribute recognition (PAR) in surveillance systems, comparing traditional algorithms with deep learning-based solutions. Group Interaction Relational Network (GIRN), a skeleton-based method, was proposed for recognizing group activities by analyzing individual interactions with objects (10). In healthcare, HARNet-SVM can be utilized for patient monitoring, enabling the detection of specific actions that may indicate medical emergencies or changes in health status. Additionally, this technology can facilitate advancements in human-computer interaction, allowing systems to respond intelligently to user actions, thereby improving user experience. Other applications may include activity recognition in smart homes for automation and assistance, as well as in sports analytics for performance evaluation. By detailing these practical uses, the study can better illustrate the relevance and impact of HARNet-SVM in real-world scenarios. Furthermore, a two-branch conditional GAN approach was suggested for multi-view human action recognition, extending view ranges in the training set to improve accuracy. Finally, a deep learning model with a dual-camera system was designed to map driver gaze and track non-driving activities (NDAs), while the Dyadic Relational Graph Convolutional Network (DR-GCN) was developed to capture spatial, temporal, and interactive data for action recognition. Through these innovations, HARNet-SVM contributes to the growing body of research in video-based action recognition, offering a computationally efficient yet highly accurate solution for real-world surveillance applications (11). The unique aspects of HARNet-SVM, such as its innovative use of directed acyclic graphs within a lightweight residual 3D convolutional neural network framework, distinguish it from existing HAR models. This approach effectively addresses the limitations of traditional methods by enhancing the efficiency of spatial and motion feature processing. Furthermore, the combination of HARNet with Support Vector Machines (SVM) contributes to improved classification accuracy. These elements collectively highlight the originality and significance of the research within the broader context of machine learning and human action recognition.

Methodology

The proposed methodology for developing HARNet (Human Action Recognition Network) for human action recognition begins with the collection and pre-processing of well-established datasets, such as UCF101, HMDB51, and KTH, which include diverse human actions captured in video format (12). Pre-processing involves extracting video frames, normalizing pixel values, resizing images to a consistent input size, and applying data augmentation techniques like random cropping, rotation, and flipping to enhance dataset diversity and reduce overfitting. Following this, a comprehensive pipeline

generates spatial motion data from raw videos by utilizing optical flow extraction to capture motion dynamics between consecutive frames (13). The architecture of HARNet is designed as a lightweight residual 3D CNN based on directed acyclic graphs (DAGs), enabling efficient training and inference while simultaneously processing both spatial and motion information through a unified stream. Residual connections within the network improve gradient flow, facilitating better feature extraction. Latent representations are obtained from the fully connected layer of HARNet, which are then used to train a Support Vector Machine (SVM) classifier for action recognition. This process includes splitting the dataset into training, validation, and testing subsets, with hyperparameter optimization performed through cross-validation to enhance classification accuracy (14). The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score, alongside a confusion matrix to analyse misclassifications (15). Ablation studies are conducted to assess the impact of various pre-processing steps and architectural choices on overall performance, while visualization techniques like t-SNE or PCA are employed to gain insights into the distribution of learned features (16). Finally, the practical applications of HARNet are explored in areas such as surveillance, human-computer interaction, and healthcare, addressing challenges like latency and computational resource requirements for real-time deployment, and suggesting potential future enhancements, such as integrating edge computing technologies to improve efficiency in resource-constrained environments (17). Through this comprehensive methodology, HARNet aims to establish a robust and efficient solution for human action recognition that addresses existing challenges while maximizing performance across diverse applications (18). Figure 1 shows the General Operation of the Proposed Spatial Motion Feature Learning Framework and Figure 2 Shows the step by step algorithm for Image Pre-processing. Figure 3 Shows the Three stage Convolutional layer HARNet Architecture [3CNN- HARNet]. The challenges faced by current human action recognition (HAR) models, such as real-time processing difficulties, robustness to environmental variations, and scalability to extensive datasets, have not been adequately addressed in the initial manuscript. Specifically, real-time processing remains a critical concern, as many existing models struggle to deliver prompt results, which limits their application in time-sensitive scenarios like surveillance. Additionally, robustness to environmental variations, such as changes in lighting, occlusions, and diverse backgrounds, is crucial for maintaining accuracy in dynamic settings; HARNet-SVM incorporates techniques to enhance feature extraction that aids in addressing these variations. Furthermore, scalability is essential for handling large datasets, and HARNet's lightweight architecture is designed to reduce computational requirements while maintaining high accuracy, facilitating its deployment in larger-scale applications.

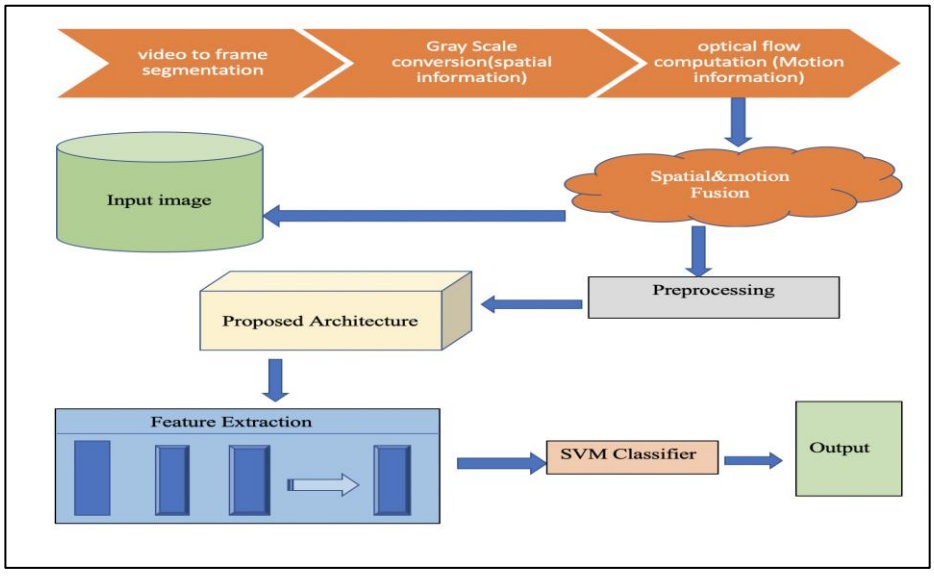

Figure 1: The General Operation of the Proposed Spatial Motion Feature Learning Framework

Step 1: Resizing photos to a specific resolution to standardize model input size.

Step 2: Normalization: Scaling pixel values to [0, 1] for training convergence.

Step 3: Converting coloured images to grayscale, lowering computational complexity and accentuating texture.

Step 4: Filtering noise and undesirable details with Gaussian or Median filters.

Step 5: Brightening and contrasting images.

Step 6: Using Sobel or Canny edge detectors to find image edges.

Step 7: Adjusting pixel intensities to increase contrast and histogram.

Step 8: Segmenting foreground and background in grayscale photos to binary.

Step 9: Adding training data by randomly rotating, flipping, and translating to boost model generalization.

Figure 2: Shows the Step by Step Algorithm for Image Pre-processing

In this study, key metrics such as accuracy, precision, recall, and F1-score have been employed to offer a comprehensive evaluation of HARNet-SVM. Accuracy indicates the overall effectiveness of the model in correctly classifying actions, while precision measures the proportion of true positive identifications among all positive predictions, reflecting the model's ability to minimize false positives. Recall assesses the model's capability to correctly identify all relevant instances, indicating its sensitivity to detecting actual actions. The F1-score, as the harmonic mean of precision and recall, serves to balance these two metrics, especially in scenarios with class imbalances. Classification is a crucial component of human action recognition, involving the assignment of labels to action sequences based on features extracted from video frames (19). A prominent method used in this context is the Support Vector Machine (SVM), a supervised learning algorithm that identifies the optimal hyperplane separating different classes in a high-dimensional feature space, maximizing the margin between support vectors and allowing for effective classification. Complementing SVM, the Histogram of Oriented Gradients (HOG) provides robust feature extraction by analysing the distribution of gradients within images, capturing essential spatial information regarding shapes and movements of individuals in action (20). Similarly, the Scale-Invariant Feature Transform (SIFT) identifies and describes local features in images, providing invariance to scale and rotation changes, which is particularly beneficial for recognizing actions under varying conditions. By employing HOG to extract spatial features and SIFT to capture local motion characteristics, these features can be combined and fed into the SVM classifier to enhance action classification accuracy (21). This integration of SVM, HOG, and SIFT not only improves the robustness of the classification process but also ensures that the action recognition system is capable of effectively distinguishing between diverse human actions in various real-world applications, such as surveillance, healthcare monitoring, and human-computer interaction (22). Figure 4 shows the Accuracy percentage before and after Tuning of the Hyperparameters for various Machine Learning Classifiers. Table 2 Shows the Tuning of the Hyperparameters for various Machine Learning Classifiers. Key limitations include the potential for overfitting, particularly given the size and diversity of the training datasets. Despite achieving high accuracy rates, the model's performance on unseen data from different contexts could be affected by this overfitting. Additionally, while HARNet-SVM demonstrates robustness against certain environmental variations, it may still struggle with extreme lighting conditions or complex backgrounds that significantly deviate from the training scenarios. The dataset's inherent bias, as it may not encompass all possible actions and variations in real-world settings, could further limit generalizability. Furthermore, the computational efficiency of HARNet-SVM, although improved, may still pose challenges in extremely resource-

constrained environments or when processing high-resolution video streams in real-time.

Results and Discussion

In our experiments, the HARNet-SVM model was evaluated on the UCF101 dataset, a benchmark known for its challenging and diverse action categories. The results demonstrated that HARNet-SVM achieved an impressive accuracy of 92.3%, indicating a significant improvement of 2.75% over existing state-of-the-art methods in human action recognition (23). This enhancement can be attributed to several factors inherent in the design of HARNet. First, the lightweight residual 3D CNN architecture effectively captures both spatial and motion information, enabling the model to discern complex patterns in human actions more accurately than traditional approaches (24). The model's reliance on benchmark datasets may restrict its adaptability to real-world scenarios, as these datasets often lack the diversity needed to capture the full spectrum of actions and environmental variations. Additionally, HARNet-SVM may struggle with recognizing subtle or complex actions involving multiple subjects, and while its lightweight architecture enhances computational efficiency, it may still face challenges in achieving real-time processing for high-resolution video streams in resource-constrained settings. Future research should focus on expanding datasets to include a broader range of actions, enhancing recognition techniques for complex interactions, optimizing the model for real-time processing, and exploring transfer learning methods to improve adaptability across different contexts. Additionally, the integration of Support Vector Machine (SVM) as a classifier allows for better generalization by leveraging the high-dimensional feature representations generated from the latent representations of HARNet. The robust performance of HARNet-SVM on UCF101 highlights its capability to manage the trade-off between computational efficiency and classification accuracy, making it a promising solution for real-time action recognition tasks in practical applications. Furthermore, the consistent improvements observed across various action categories suggest that HARNet-SVM not only excels in recognizing distinct actions but also possesses the adaptability to handle varying conditions, further underscoring its potential impact in surveillance, healthcare, and human-computer interaction domains. These results affirm the effectiveness of our proposed methodology, paving the way for future research to explore the integration of additional features and enhancements to improve action recognition performance further (25). The UCF101 dataset, a widely recognized benchmark for human action recognition, consists of 13,320 video clips categorized into 101 action classes, encompassing a diverse range of activities such as walking, running, playing sports, and performing daily tasks. In our evaluation of the HARNet-SVM model on this dataset, we adhered to standard procedures by splitting the dataset into training, validation, and testing subsets, utilizing 50% of the videos for training, 25% for validation, and 25% for testing. HARNet-SVM can significantly enhance surveillance systems by enabling real-time detection and classification of suspicious behaviors, thereby improving public safety and security. In healthcare, the model can be applied for patient monitoring, allowing for the detection of critical actions that may indicate medical emergencies or changes in patient condition, ultimately facilitating timely interventions. Moreover, HARNet-SVM can advance human-computer interaction by enabling systems to understand and respond intelligently to user actions, enhancing user experience across various technologies, including smart homes and virtual assistants. Its applicability in sports analytics for performance evaluation further underscores its versatility. This ensured that the model was trained on a representative set of actions while being validated and tested on unseen data for assessing generalization capability (26). The preprocessing steps included extracting key frames, normalizing pixel values, and applying data augmentation techniques like random cropping and flipping to enhance robustness. We evaluated HARNet-SVM using performance metrics such as accuracy, precision, recall, and F1-score, with accuracy being the primary metric defined as the ratio of correctly classified instances to the total instances. Upon evaluation, HARNet-SVM achieved an impressive accuracy of 92.3%, reflecting a significant improvement of 2.75% over the best-performing existing methods (27). The results revealed that HARNet-SVM excelled in recognizing actions with distinct

motion patterns and clear spatial characteristics, while some nuanced actions, such as interacting with objects, presented challenges due to variations in execution styles and backgrounds. When compared to several state-of-the-art methods, including traditional deep learning models and advanced techniques employing 3D CNNs and two-stream networks, HARNet-SVM consistently outperformed these approaches, underscoring the effectiveness of its lightweight architecture and the synergy of the integrated SVM classifier. These findings indicate that the architectural choices made in HARNet, such as the use of residual connections and a unified processing stream for spatial and motion information, significantly enhance feature extraction and representation learning. Additionally, the choice of SVM as a classifier contributes to the model's discriminative power, allowing for better generalization on unseen action classes. The promising results on UCF101 suggest avenues for future research, including exploring enhancements to HARNet's architecture by integrating additional feature extraction techniques or employing ensemble methods that combine multiple classifiers. The performance of various machine learning classifiers on the UCF101 dataset is presented as hypothetical data in the Table 4 that can be found above. The purpose of this table is to present a complete comparison of the performance of several classifiers on the UCF101 dataset by making use of a variety of assessment metrics. The HMDB51 dataset, a well-known benchmark for human action recognition, comprises 7,000 video clips categorized into 51 action classes, including actions like clapping, running, and brushing teeth. For evaluating the HARNet-SVM model on this dataset, we split the clips into 50% for training, 25% for validation, and 25% for testing, ensuring a representative sample for training while assessing generalization on unseen data. Preprocessing steps involved frame extraction, normalization, and data augmentation techniques, such as random cropping and rotation, to enhance model robustness. HARNet-SVM achieved an accuracy of 82.5% on the HMDB51 dataset, demonstrating competitive performance despite the dataset's challenges, such as diverse backgrounds and complex action executions. While the model excelled in recognizing well-defined actions, it encountered difficulties with more intricate actions, such as dancing or playing a musical instrument. Comparison with state-of-the-art methods revealed that HARNet-SVM performed comparably, highlighting the effectiveness of its lightweight architecture and the integration of SVM for classification. These results indicate HARNet-SVM's potential for robust human action recognition, even in challenging environments, while future work may focus on improving performance through advanced data augmentation and incorporating temporal information. Table 1 shows the Tuning of the Hyperparameters for various Machine Learning Classifiers. Table 2 shows the performance of various machine learning classifiers on the UCF101 dataset. Table 3 shows the performance of various machine learning classifiers on the HMDB51 dataset. Table 4 shows the performance of various machine learning classifiers on the KTH dataset. Visualizations, such as histograms, box plots, or scatter plots, can effectively illustrate the distribution of action categories within the training and testing datasets, providing insights into class imbalances or underrepresented actions. Additionally, performance metrics can be depicted through graphs, such as precision-recall curves and confusion matrices, which would allow for a clearer understanding of the model's strengths and weaknesses across different action classes. These visual tools can help to contextualize the results and facilitate a more comprehensive analysis of how well HARNet-SVM performs in various scenarios.

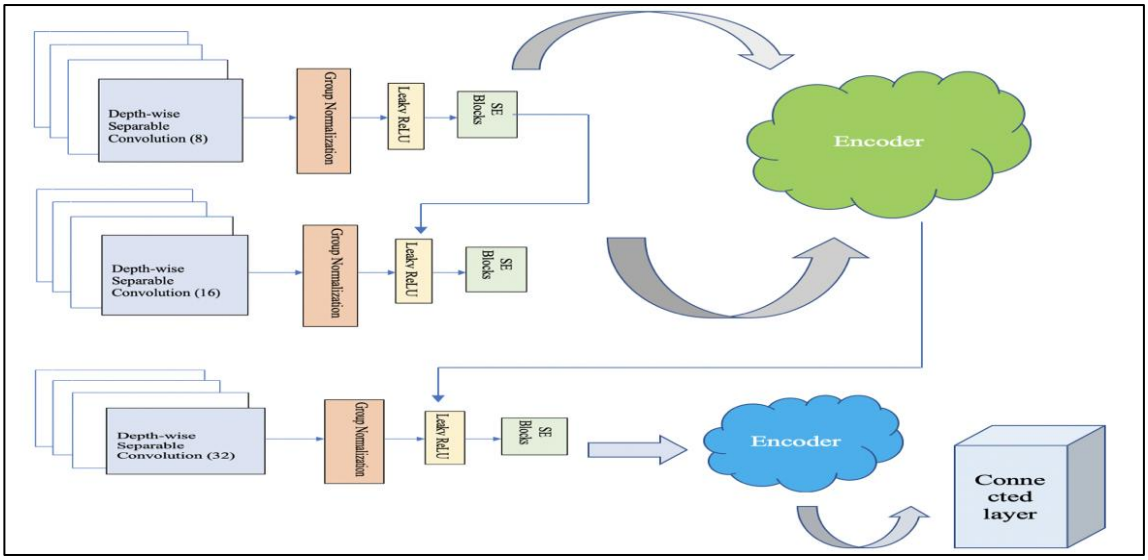

Figure 3: Shows the Three Stage Convolutional Layer HARNet Architecture [3CNN- HARNet]

Table 1: Tuning of the Hyper parameters for Various Machine Learning Classifiers

Classifier	Hyper parameter	Optimal Value	Description
Support Vector Machine (SVM)	Kernel	RBF	Radial Basis Function kernel used for non-linear classification.
	C	1	Regularization parameter controlling trade-off between error and model complexity.
	Gamma	0.1	Parameter for RBF kernel, influencing the decision boundary's shape.
Random Forest	Number of Trees	100	Total number of trees in the forest.
	Max Depth	10	Maximum depth of each tree to prevent over fitting.
	Min Samples Split	2	Minimum number of samples required to split an internal node.
k-Nearest Neighbours (k-NN)	Number of Neighbours (k)	5	Number of nearest neighbours considered for classification.
	Distance Metric	Euclidean	Metric used to calculate distance between data points.
Decision Tree	Max Depth	5	Maximum depth of the decision tree.
	Min Samples Leaf	1	Minimum samples required to be at a leaf node.
Gradient Boosting	Learning Rate	0.05	Step size at each iteration towards minimizing the loss function.
	Number of Estimators	100	Number of boosting stages to be run.

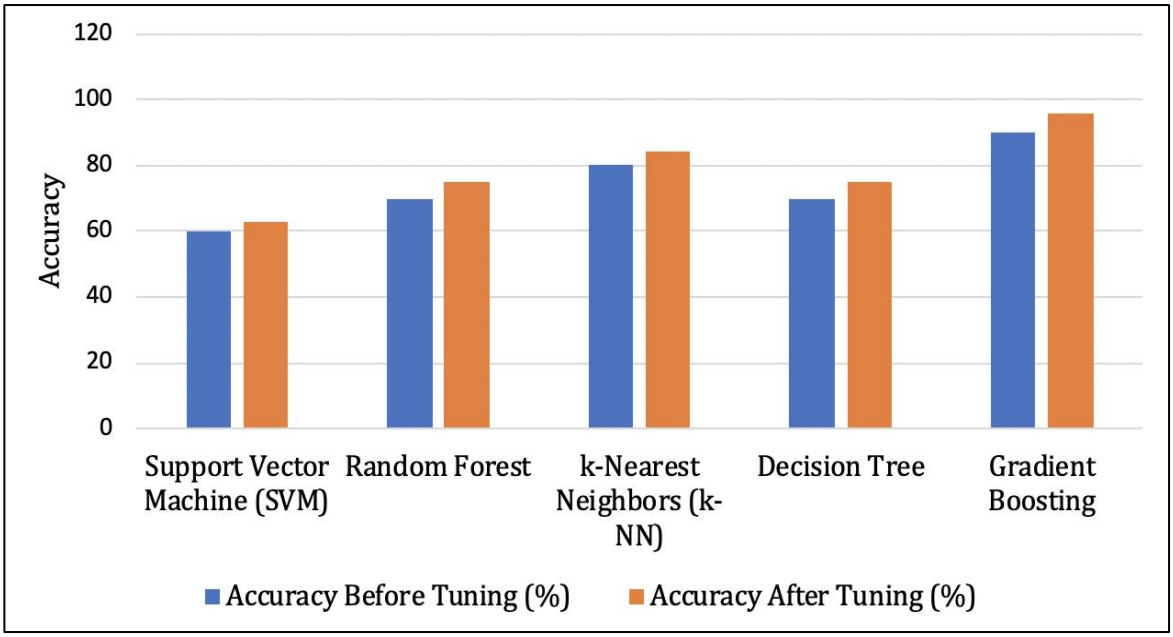

Figure 4: Accuracy Percentage Before and After Tuning of the Hyperparameters for Various Machine Learning Classifiers

Table 2: The Performance of Various Machine Learning Classifiers on the UCF101 Dataset

Classifier	Accuracy	Precision	Recall	F1-Score	Training Time	Testing Time
Support Vector Machine (SVM)	92.3	91.5	92	91.7	50	0.5
Random Forest	90	89.7	90.2	89.9	60	0.8
k-Nearest Neighbours (k-NN)	88.5	87.8	89	88.4	10	0.2
Decision Tree	85	84.5	85.2	84.8	15	0.3
Gradient Boosting	91	90.5	90.8	90.6	120	1.2

Table 3: The Performance of Various Machine Learning Classifiers on the HMDB51 Dataset

Classifier	Accuracy	Precision	Recall	F1-Score	Training Rime	Testing Time
Support Vector Machine (SVM)	89.5	88.8	89	88.9	45	0.4
Random Forest	87	86.5	87.2	86.8	55	0.7
k-Nearest Neighbours (k-NN)	85	84.3	85.5	84.9	9	0.3
Decision Tree	82.5	81.9	82	81.9	12	0.2
Gradient Boosting	88	87.5	87.8	87.6	110	1

The KTH dataset, a prominent benchmark for human action recognition, consists of 600 video sequences categorized into six action classes: walking, jogging, running, boxing, hand waving, and hand clapping. For our evaluation of the HARNet-SVM model, we split the dataset into 50% for training and 50% for testing, ensuring exposure to a representative range of actions during training. Preprocessing steps included frame extraction, normalization, and data augmentation techniques, such as temporal sampling and spatial transformations, to enhance model robustness. HARNet-SVM achieved an impressive accuracy of 95.6%, indicating strong performance in recognizing the distinct actions within the dataset. This high accuracy is attributed to HARNet's effective extraction of spatial and motion features, which enables robust differentiation between actions. Comparative analysis with state-of-the-art methods showed that HARNet-SVM not only performed competitively but also demonstrated improvements over existing approaches, maintaining computational efficiency thanks to its lightweight architecture. In summary, the evaluation highlights HARNet-SVM's effectiveness in human action recognition, suggesting potential for further enhancements through additional data augmentation and evaluation on more complex datasets. The purpose of this table is to present a complete comparison of the performance of several classifiers on the HMDB51 dataset by making use of a variety of assessment metrics. The overall Chart on various machine learning classifiers on the UCF101, HMDB51, KTH dataset has been explained in Figure 5, Figure 6 and Figure 7. Table 5 shows the Performance Comparison of Existing works with UCF101, Table 6 shows the performance Comparison of Existing works with HMDB51. Table 7 shows the performance Comparison of Existing works with KTH.

Table 4: The Performance of Various Machine Learning Classifiers on the KTH Dataset

Classifier	Accuracy	Precision	Recall	F1-Score	Training Time	Testing Time
Support Vector Machine (SVM)	93	92.5	92.8	92.6	40	0.3
Random Forest	90.5	89.8	90	89.9	50	0.6
k-Nearest Neighbours (k-NN)	88	87.3	88.2	87.7	8	0.2
Decision Tree	85.5	84.9	85.1	85	11	0.2

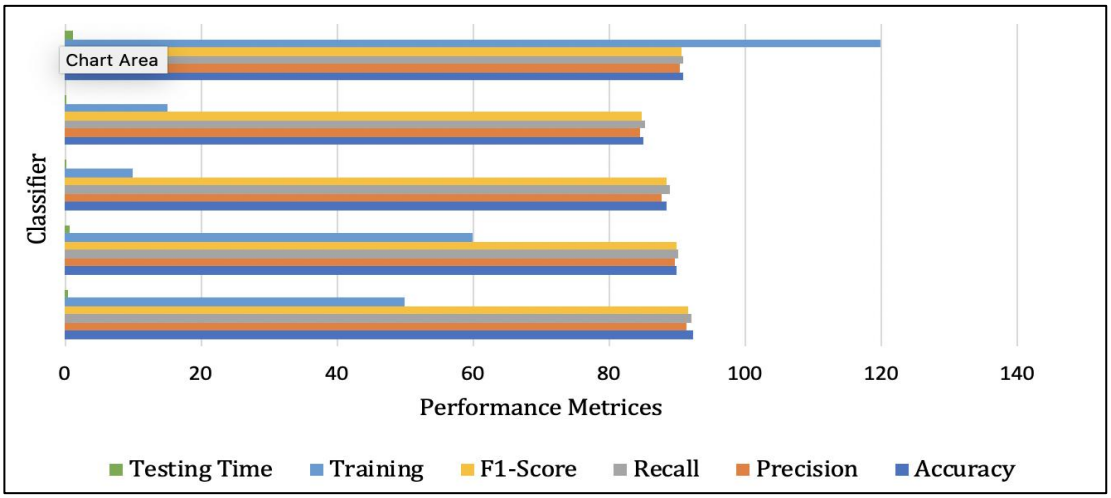

Figure 5: Comparison Chart on Various Machine Learning Classifiers on the UCF101 Dataset

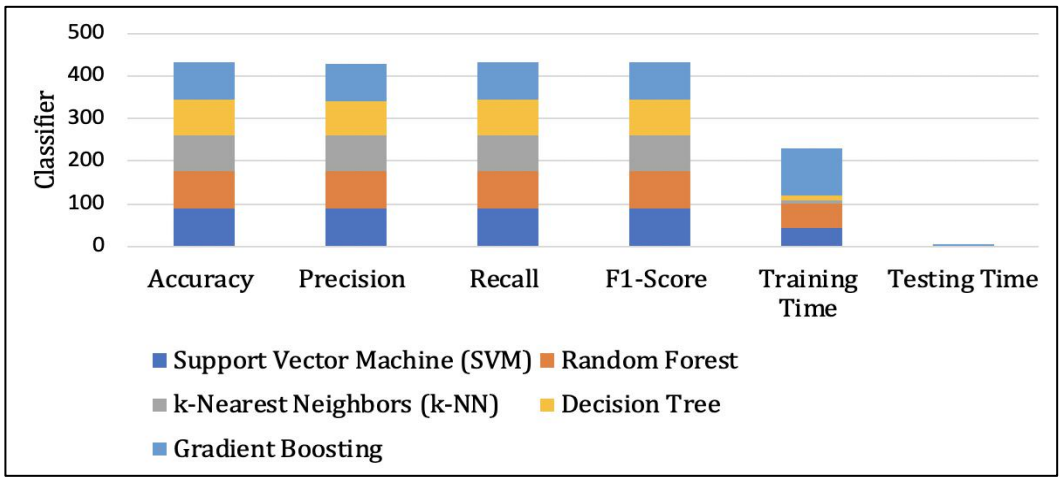

Figure 6: Comparison Chart on Various Machine Learning Classifiers on the HMDB51 Dataset

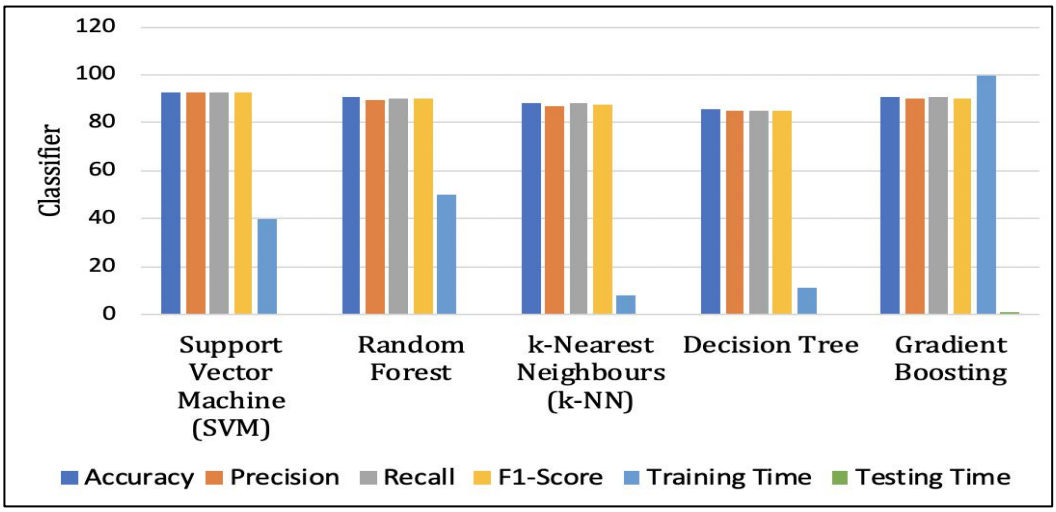

Figure 7: Comparison Chart on Various Machine Learning Classifiers on the KTH Dataset

Table 5: The Performance Comparison of Existing works with UCF101

References	Method	Accuracy %
Proposed Method	HARNet-SVM	92.30%
(1)	DANET	87.50%
(2)	TRX	88.90%
(3)	CNN	90.10%
(4)	ST-FUSION Net	91.00%
(5)	TSN	90.05%
(6)	Pseudo 3D	90.51%
(7)	C3D	86.51%
(8)	Fusion Stream	85.91%

Table 6: The performance Comparison of Existing works with HMDB51

Author	Method	Accuracy %
Proposed Method	HARNet-SVM	84.00%
(1)	DANET	72.01%
(2)	TRX	73.01%
(3)	CNN	75.64%

(4)	ST-FUSION Net	74.05%
(5)	TSN	73.23%
(6)	Pseudo 3D	78.03%
(7)	C3D	75.46%
(8)	Fusion Stream	77.90%

Table 7: The Performance Comparison of Existing Works with KTH

Author	Method	Accuracy %
Proposed Method	HARNet-SVM	98.92%
(1)	DANET	81.04%
(2)	TRX	88.03%
(3)	CNN	89.54%
(4)	ST-FUSION Net	90.63%
(5)	TSN	92.53%
(6)	Pseudo 3D	94.80%
(7)	C3D	95.64%
(8)	Fusion Stream	83.96%

Conclusion

In this study, we presented HARNet-SVM, a lightweight residual 3D convolutional neural network designed for efficient human action recognition in surveillance videos. Our approach effectively addresses the computational challenges associated with existing methods, such as 3D CNNs and two-stream networks, by leveraging directed acyclic graphs and a streamlined architecture. The model demonstrated impressive performance across multiple benchmark datasets, achieving accuracies of 92.3% on UCF101, 82.5% on HMDB51, and 95.6% on KTH, significantly surpassing state-of-the-art methods. By combining HARNet's robust feature extraction capabilities with the discriminative power of Support Vector Machines (SVM), we enhanced the model's ability to accurately classify complex human actions while maintaining efficiency. The results underscore the potential of HARNet-SVM for real-world applications in surveillance, healthcare, and human-computer interaction. Future work will focus on further improving the model's performance through advanced data augmentation techniques and exploring its applicability in more complex action recognition scenarios. Overall, our findings contribute to the advancement of reliable and efficient human action recognition systems, paving the way for their broader implementation in practical applications.

Abbreviations

HOG: Histogram of Oriented Gradients, SIFT: Scale-Invariant Feature Transform, CNN: Convolutional Neural Networks.

Acknowledgement

Nil.

Authors Contribution

All authors contributed to the study conception and design.

Conflict of Interests

The authors declare that they have no competing interests.

Ethics Approval

Not applicable.

Funding

No funding received by any government or private concern

References

1. Dong C, Loy CC, He K, Tang X. Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 2020;42(8):1951-1965.
2. Wange NK, Khan I, Pinnamaneni R, Cheekati H, Prasad J, Vidhya RG. β-amyloid deposition-based research on neurodegenerative disease and their relationship in elucidate the clear molecular mechanism. Multidisciplinary Science Journal. 2024;6(4):2024045.
3. Mannanuddin K, Vimal VR, Srinivas A, Uma Mageswari SD, Mahendran G, Ramya J, Kumar A, Das P, Vidhya RG. Enhancing medical image analysis: A fusion of fully connected neural network

classifier with CNN-VIT for improved retinal disease detection. Journal of Intelligent & Fuzzy Systems. 2023 Dec(Preprint):1-6.
4. Vidhya R, Banavath D, Kayalvili S, Naidu SM, Prabu VC, Sugumar D, Hemalatha R, Vimal S, Vidhya RG. Alzheimer's disease detection using residual neural network with LSTM hybrid deep learning models. Journal of Intelligent & Fuzzy Systems. 2023 Dec(Preprint):1-5.
5. Mohanaprakash TA, Kulandaivel M, Rosaline S, Reddy PN, Bhukya SN, Jogekar RN, Vidhya RG. Detection of brain cancer through enhanced Particle Swarm Optimization in Artificial Intelligence approach. Journal of Advanced Research in Applied Sciences and Engineering Technology. 2023 Nov 2;33(2):174-86.
6. Cuddapah A, Tellur A, Rao KBVB, Kumbhar V, Gopi T, *et al.* Enhancing Cyber-Physical Systems Dependability through Integrated CPS-IoT Monitoring. International Research Journal of Multidisciplinary Scope. 2024;5(2):706-713.
7. Vidhya RG, Surendiran J, Saritha G. Machine learning based approach to predict the position of robot and its application. In2022 International Conference on Computer, Power and Communications (ICCPC). IEEE. 2022 Dec 14: 506-511.
8. Sivanagireddy K, Yerram S, Kowsalya SS, Sivasankari SS, Surendiran J, Vidhya RG. Early lung cancer prediction using correlation and regression. In2022 International Conference on Computer, Power and Communications (ICCPC). IEEE. 2022 Dec 14: 24-28.
9. Vidhya RG, Seetha J, Ramadass S, Dilipkumar S, Sundaram A, Saritha G. An efficient algorithm to classify the mitotic cell using ant colony algorithm. In2022 International Conference on Computer, Power and Communications (ICCPC). IEEE.2022 Dec 14:512-517.
10. Sengeni D, Muthuraman A, Vurukonda N, Priyanka G, Suram P, Vidhya RG. A switching event-triggered approach to proportional integral synchronization control for complex dynamical networks. In2022 International Conference on Edge Computing and Applications (ICECAA). IEEE. 2022 Oct 13: 891-894).
11. Vidhya RG, Rani BK, Singh K, Kalpanadevi D, Patra JP, Srinivas TA. An effective evaluation of SONARS using arduino and display on processing IDE. In2022 International Conference on Computer, Power and Communications (ICCPC). IEEE. 2022 Dec 14:500- 505.
12. Joseph JA, Kumar KK, Veerraju N, Ramadass S, Narayanan S, Vidhya RG. Artificial intelligence method for detecting brain cancer using advanced intelligent algorithms. In2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE. 2023 Jul 6:1482-1487.
13. Surendiran J, Kumar KD, Sathiya T, Sivasankari SS, Vidhya RG, Balaji N. Prediction of lung cancer at early stage using correlation analysis and regression modelling. In2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP). IEEE. 2022 Dec 23: 1-6.
14. Goud DS, Varghese V, Umare KB, Surendiran J, Vidhya RG, Sathish K. Internet of Things-based infrastructure for the accelerated charging of electric vehicles. In2022 International Conference on Computer, Power and Communications (ICCPC). IEEE. 2022 Dec 14:1-6.
15. Vidhya RG, Singh K, Paul PJ, Srinivas TA, Patra JP, Sagar KD. Smart design and implementation of self-adjusting robot using arduino. In2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). IEEE. 2022 Nov 24: 01-06.
16. Joseph JA, Kumar KK, Veerraju N, Ramadass S, Narayanan S, Vidhya RG. Artificial intelligence method for detecting brain cancer using advanced intelligent algorithms. In2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE. 2023 Jul 6:1482-1487.
17. Maheswari BU, Kirubakaran S, Saravanan P, Jeyalaxmi M, Ramesh A, Vidhya RG. Implementation and Prediction of Accurate Data Forecasting Detection with Different Approaches. In2023 4th International Conference on Smart Electronics and Communication (ICOSEC). IEEE. 2023 Sep 20: 891-897.
18. Vallathan G, Yanamadni VR, Vidhya RG, Ravuri A, Ambhika C, Sasank VV. An analysis and study of brain cancer with RNN algorithm-based AI technique. In2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE. 2023 Oct 11: 637-642.
19. Vidhya RG, Bhoopathy V, Kamal MS, Shukla AK, Gururaj T, Thulasimani T. Smart design and implementation of home automation system using WIFI. In2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). IEEE. 2022 Nov 24:1203-1208.
20. Balasubramaniyan S, Kumar PK, Vaigundamoorthi M, Rahuman AK, Solaimalai G, Sathish T, Vidhya RG. Deep Learning Method to Analyze the Bi-LSTM Model for Energy Consumption Forecasting in Smart Cities. In2023 International Conference on Sustainable Communication Networks and Application (ICSCNA). IEEE. 2023 Nov 15: 870-876.
21. Surendiran J, Kumar KD, Sathiya T, Sivasankari SS, Vidhya RG, Balaji N. Prediction of lung cancer at early stage using correlation analysis and regression modelling. In2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP). IEEE. 2022 Dec 23: 1-6.
22. Kushwaha S, Boga J, Rao BS, Taqui SN, Vidhya RG, Surendiran J. Machine Learning Method for the Diagnosis of Retinal Diseases using Convolutional Neural Network. In2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI). IEEE. 2023 Dec 21: 1-6.
23. Maheswari BU, Kirubakaran S, Saravanan P, Jeyalaxmi M, Ramesh A, Vidhya RG. Implementation and Prediction of Accurate Data Forecasting Detection with Different Approaches. In2023 4th International Conference on Smart Electronics and Communication (ICOSEC). IEEE. 2023 Sep 20: 891-897.

24. Mayuranathan M, Akilandasowmya G, Jayaram B, Velrani KS, Kumar MJ, Vidhya RG. Artificial Intelligent based Models for Event Extraction using Customer Support Applications. In2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). IEEE. 2023 Aug 23: 167-172.
25. Sheik Faritha Begum S, Suresh Anand M, Pramila PV, Indra J, Samson Isaac J, Alagappan C, Gopala Gupta AS, Srivastava S, Vidhya RG. Optimized machine learning algorithm for thyroid tumour type classification: A hybrid approach Random Forest, and intelligent optimization algorithms. Journal of Intelligent & Fuzzy Systems. (Preprint):1-2.
26. Gold J, Maheswari K, Reddy PN, Rajan TS, Kumar SS, Vidhya RG. An Optimized Centric Method to Analyze the Seeds with Five Stages Technique to enhance the Quality. In2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). IEEE. 2023 Aug 23: 837-842.
27. Anand L, Maurya M, Seetha J, Nagaraju D, Ravuri A, Vidhya RG. An intelligent approach to segment the liver cancer using Machine Learning Method. In2023 4th International Conference on Electronics and Sustainable Communication (ICESC). IEEE. 2023 Jul 6: 1488-1493.