ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/

Deep Learning and Bat Algorithm-Based Human Activity Recognition

V. Rojas[1,*]

[1] Department of Electrical Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

ARTICLE INFO

Article History:
Received 5 April 2025
Received in revised form 11 July 2025
Accepted 22 September 2025
Available online 28 September 2025

Keywords:
Human Activity Recognition, Deep Learning, Convolutional Neural Networks, Bat Optimization Algorithm, Intelligent Surveillance, Weizmann Dataset

ABSTRACT

Human Activity Recognition (HAR) is a rapidly evolving research area that focuses on identifying and classifying human activities using sensor- and vision-based data. HAR has gained significant attention due to its wide range of applications in intelligent surveillance systems, biometric identification, smart environments, healthcare monitoring, and human–computer interaction. Accurate and real-time activity recognition is particularly important in surveillance applications, where the timely detection of suspicious behaviors can contribute to crime prevention and public safety. Recent advances in deep learning have demonstrated remarkable performance in HAR tasks, especially through Convolutional Neural Networks (CNNs), which are capable of automatically extracting discriminative features from visual data. However, the performance of CNN-based models is highly dependent on the selection of optimal network parameters. To address this challenge, this paper proposes a hybrid HAR framework that integrates CNNs with the Bat Optimization Algorithm (BOA) to enhance feature learning and classification performance. The proposed method is evaluated using the Weizmann human activity dataset and compared with several existing HAR approaches. Experimental results demonstrate that the integration of BOA with CNN improves recognition accuracy and classification effectiveness, highlighting the potential of the proposed framework for intelligent activity recognition applications.

1. INTRODUCTION

Human Activity Recognition (HAR) is a pivotal domain within data science, artificial intelligence, and signal processing aimed at identifying and classifying human actions based on sensory data, images, and video sequences. This technology facilitates a deeper comprehension of how humans interact with their environment and digital interfaces, which has garnered significant research attention in recent years [1-3].

The significance of HAR lies in its ability to endow intelligent frameworks with human behavioral understanding. For instance, a security system will not merely record video feeds but will actively recognize suspicious activities, such as physical altercations or theft. In healthcare, precise tracking of body kinetics can monitor patients in real time and dispatch automated alerts during critical emergencies. This capability enhances public safety, elevates the quality of life, and streamlines human-machine interaction.

* Corresponding Author: ValentinaRojas@gmail.com
Department of Electrical Egineering, Pontificia Universidad Católica de Chile, Santiago, Chile

http://dx.doi.org/10.22034/tmi.2025.244864

In the security sector, HAR serves as an intelligent layer for conventional surveillance cameras. Frameworks driven by this technology can identify anomalous activities such as unauthorized entry, physical assaults, or erratic movements and immediately notify the monitoring operator, thereby minimizing response times and preempting criminal acts. Additionally, HAR aids in crowd management and traffic control by analyzing collective group dynamics.

Another major application area for HAR is in healthcare and remote clinical monitoring. For example, wearable systems integrated with inertial sensors can identify physical actions such as walking, running, or falling, and subsequently relay the patient's or elderly individual's structural status to medical professionals or family members. Furthermore, in smart homes, HAR enables device control based on body movements or hand gestures, fostering a more seamless and seamless user experience [4-7].

Despite these advancements, several challenges persist in this domain. The intrinsic intra-class variability of human actions, environmental fluctuations (such as lighting conditions or camera viewpoints), and inter-subject physiological differences frequently degrade system accuracy. Moreover, training robust deep learning architectures requires massive, accurately labeled datasets, the curation of which is both costly and time-consuming. User privacy also introduces prominent ethical concerns in surveillance deployment.

Deep learning has managed to mitigate a substantial portion of these challenges. Convolutional Neural Networks (CNNs) possess the capacity to automatically extract spatial image and motion features, while Recurrent Neural Networks (RNNs) such as LSTM and GRU are highly effective at modeling sequential temporal variations. Fusing these two paradigms allows for high-fidelity activity recognition even within complex environments. Additionally, intelligent metaheuristic optimization algorithms such as the Bat Algorithm, Genetic Algorithm, or Particle Swarm Optimization (PSO) can optimize the model's training process, culminating in superior classification accuracy [8-10].

2. EXISTING METHODS

In study [1], human activity recognition is executed using the Weizmann dataset and Convolutional Neural Networks (CNNs). Three distinct deep architectures namely VGG-16, VGG-19, and Google’s InceptionNet-v3 are deployed to classify the target activities. Additionally, a transfer learning paradigm is exploited to leverage pre-trained knowledge from large-scale repositories such as ImageNet. In the initial phase, consecutive video frames corresponding to each activity are extracted, after which visual feature descriptors are extracted via transfer learning and channeled into the classification pipeline. Finally, the performances of the different CNN models are benchmarked, and the empirical results are validated against state-of-the-art studies utilizing identical datasets.

In reference [2], a novel deep framework based on the Long Short-Term Memory (LSTM) network is introduced, designed by hybridizing the operational capabilities of Recurrent Neural Networks (RNNs) and CNNs. This model is deployed within smart home environments to forecast subsequent sequence events. In recent years, technological breakthroughs in sensor instrumentation and communication protocols have driven exponential growth in smart home applications. To benchmark its operational efficiency, the performance of the model was validated on a real-world dataset. The empirical profiles demonstrated that LSTM networks yield superior performance in next-event prediction compared to alternative baseline methods.

In paper [3], a hybrid deep neural network framework is engineered, leveraging multiple network topologies to facilitate robust feature learning from sensor-derived data for human activity recognition. The architecture integrates convolutional layers, Bidirectional LSTM (BiLSTM) layers, and complementary auxiliary modules.

In paper [4], the problem of human activity recognition is analyzed via deep learning by harnessing features derived from multi-spectrogram representations. In this approach, three distinct spectrograms are fused together and chaneled as a unified multi-channel spectrogram into a VGG-16 deep neural network to perform the activity recognition task.

In paper [5], a model is introduced that is capable of effectively classifying six distinct user activities. This framework utilizes transfer learning initialized with a 1D Convolutional Neural Network (1D-CNN). Although convolutional architectures are primarily celebrated for their extraordinary success in processing 2D grid data and imagery, their utility is not mathematically restricted to computer vision. This topology can be seamlessly extended to time-series sequential data. For instance, numerous studies have verified that coupling a CNN with an LSTM network is highly effective for video processing; within this framework, the CNN handles spatial feature extraction, while the LSTM is responsible for sequential temporal pattern analysis.

In paper [6], a methodology for human activity classification based on 3D skeleton data is presented. In this approach, the skeleton coordinates are initially normalized and mapped into two distinct representations. These structural representations are subsequently fed in parallel into a deep network comprising convolutional and dense layers to execute concurrent feature extraction and activity classification.

In paper [7], a two-stage learning pipeline for human activity recognition is proposed utilizing accelerometer and gyroscope telemetry data. In the first stage, a binary Random Forest (RF) classifier stratifies the activities into static and dynamic macro-classes. In the second stage, a Support Vector Machine (SVM) is deployed to perform fine-grained classification of static activities, while a deep learning model based on a 1D-CNN is exploited to resolve the dynamic activities.

3. PROPOSED METHODOLOGY

Due to their robust capacity for processing complex high-dimensional data, minimizing optimization errors, and boosting inference speed and accuracy, Convolutional Neural Networks have found widespread application across industrial, clinical, and security domains. Driven by the exponential growth of available big data and hardware acceleration breakthroughs, the architectural role of CNNs in human activity recognition is expanding continuously. The overall framework of the proposed methodology is presented below (Figure 1).

3.1. Deep Learning

Deep learning is a specialized subfield of machine learning focused on processing data structures through multi-layered artificial neural networks. Modeled abstractly after biological neural networks, this paradigm is highly capable of solving complex, high-dimensional representation problems using massive datasets. Deep neural networks exploit hierarchical architectures to automatically extract and learn feature abstractions directly from raw data. Enhanced computational power and access to comprehensive benchmark repositories have propelled deep learning to become a foundational instrument in artificial intelligence, image processing, computer vision, and natural language processing (NLP) [10-15].

The technical significance of utilizing Convolutional Neural Networks for human activity recognition encompasses the following advantages:

High Activity Classification Accuracy: CNNs excel at extracting intricate, highly non-linear feature abstractions from data tensors, significantly elevating classification accuracy across diverse activity profiles.

Automated Feature Engineering: Unlike classical machine learning workflows that necessitate hand-crafted, manual feature extraction, a CNN automatically learns discriminative representations directly from raw input signals or video frames.

Efficiency in High-Dimensional and Complex Data: Convolutional topologies possess inherent structural efficiency for processing large-scale visual data and multi-variate temporal series.

Widespread Industrial and Clinical Deployment: Implementing CNNs in automated video surveillance, smart homes, and tele-health monitoring dramatically enhances system throughput, latency, and operational reliability.

Robust Generalization and Flexibility: By learning from massive datasets, these networks exhibit strong generalization capabilities when processing entirely unseen or out-of-distribution evaluation data.

The fundamental constituent layers of a standard Convolutional Neural Network include:

Convolutional Layer: This layer applies a bank of mathematical filters (kernels) to slide across the input tensor, extracting localized features such as spatial edges, textures, and motion patterns.

Activation Layer: Introduces non-linearity into the network topology (typically via functions like ReLU), allowing the deep architecture to learn complex, non-linear function approximations.

Pooling Layer: Executes down sampling operations to reduce the spatial dimensions of the feature maps, effectively maintaining translation invariance while preserving core structural descriptors and reducing computational load.

Fully Connected (Dense) Layer: Aggregates the high-level feature maps extracted by the preceding layers to perform global decision-making and final class probability mapping.

3.2. Bat Algorithm (BA)

The Bat Algorithm is an evolutionary, swarm intelligence-based metaheuristic optimization algorithm. It was developed by mimicking the foraging behavior of microbats and their unique capacity to navigate and detect prey using echolocation. The core concept of the metaheuristic is that each individual bat represents a candidate solution vector within the problem search space. By dynamically adjusting its position and velocity vector while traversing this space, the swarm iteratively converges toward the global optimum [16-17].

Within the operational loop of the Bat Algorithm, each bat is characterized by a dynamic frequency, velocity, and position vector, which are updated iteratively. The frequency parameter governs the step size and spatial exploration reach of the bat, velocity dictates the rate of position modification, and the position coordinates define the current candidate solution in the objective landscape.

Furthermore, two critical parameters namely loudness (A) and pulse emission rate (r) regulate the behavioral dynamics of the swarm. The loudness typically decays mathematically as the individual approaches an optimal solution, whereas the pulse emission rate increases progressively. This dynamic interplay preserves a strict balance between global exploration and local exploitation.

Due to its mathematical simplicity, rapid convergence characteristics, and exceptional flexibility in solving continuous and combinatorial optimization problems, the Bat Algorithm has been extensively adopted across engineering, machine learning, network optimization, and image processing. Moreover, it can be hybridized with alternative heuristics such as the Genetic Algorithm (GA) or Particle Swarm Optimization (PSO) to boost its accuracy in hyper-complex optimization landscapes.

A major advantage of the Bat Algorithm is its intrinsic capacity to maintain a fine-tuned trade-off between broad search-space exploration and localized solution exploitation. This characteristic prevents the population from stagnating in sub-optimal local minima, significantly increasing the probability of discovering the global maximum or minimum. Consequently, this algorithm has attracted substantial interest for resolving challenging, real-world engineering optimization, industrial scheduling, and intelligent system design problems.

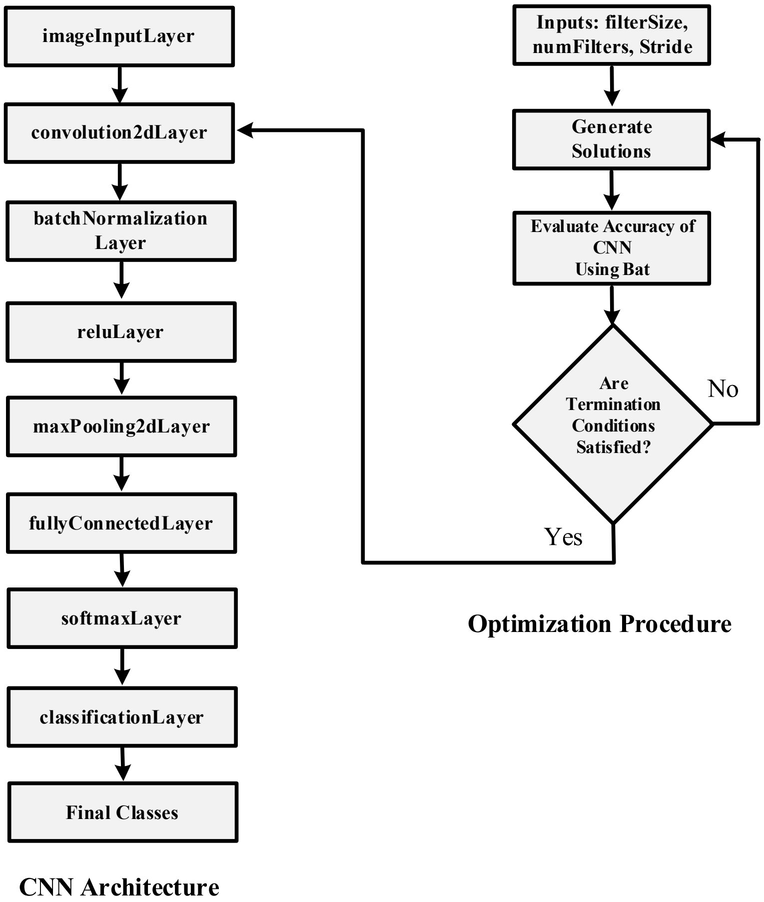

Fig. 1. Flowchart of the proposed methodology.

4. SIMULATION RESULTS

The proposed methodology and the existing comparative techniques were implemented within the MATLAB software environment developed by MathWorks. The hardware configuration utilized for the simulations consisted of an Intel(R) Core™ i5-4460 CPU @ 3.2 GHz processor equipped with 16 GB of Random Access Memory (RAM). The initialized operational parameters for the Bat Optimization Algorithm are compiled in Table 1.

Table 1. Parameters of the Bat Optimization Algorithm.

Parameter Name	Description	Value
r0	Initial Pulse Emission Rate	0.5
∝	Positive Constant 1	0.99
γ	Positive Constant 2	0.01
Population	Population Size	20
Iteration	Maximum Iterations	100

These parameters were tuned to establish an optimal balance between the convergence speed and the classification precision of the algorithm. The Weizmann dataset is a widely recognized benchmark repository within machine learning and image processing, extensively deployed for research involving human activity recognition in images and video sequences. Specifically engineered for the evaluation and simulation of human actions, this dataset features video sequences capturing 10 distinct types of activities, each performed by 9 different subjects. A representative sample of the frames from this dataset is illustrated in Figure 2 [10]. The feature map output generated by a constituent convolutional layer is displayed in Figure 3.

Fig. 2. Representative frame samples from the Weizmann dataset.

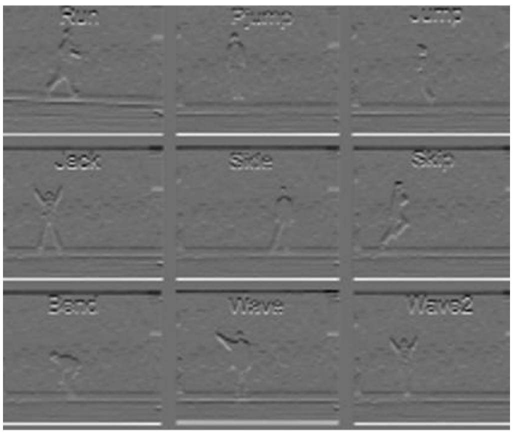

Fig. 3. Feature map output generated by a convolutional layer.

Utilizing the fully trained convolutional neural network, the predicted class labels for the Weizmann test dataset align with the ground-truth labels with an outstanding classification accuracy of 99.5%. A performance comparison benchmarking the proposed methodology against alternative state-of-the-art approaches is illustrated in Figure 4, demonstrating the superior efficiency of the presented framework. The vertical axis of the plot quantifies the classification accuracy achieved by the proposed method relative to baseline techniques.

The principal architectural advantages of the proposed framework include the automated feature learning of the convolutional neural network coupled with global parameter optimization driven by the Bat Algorithm. Conversely, in conventional methods, network hyperparameters are typically tuned empirically through trial-and-error a tedious, heuristic process that severely complicates the attainment of optimal accuracy.

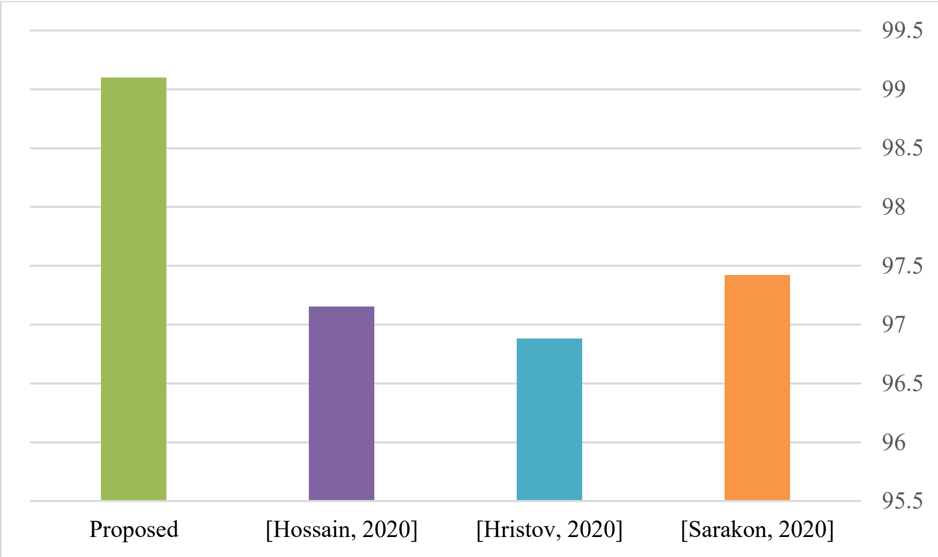

Fig. 4. Comparative accuracy analysis between the proposed method and existing methods.

5. CONCLUSION

Convolutional Neural Networks play an instrumental role in recognizing human activities from images and video sequences due to their robust capacity to automatically extract complex spatial feature hierarchies from visual data structures. Especially when classifying actions with highly non-linear and intricate motion trajectories, the deployment of a CNN dramatically improves overall recognition accuracy. In the proposed methodology, critical structural hyperparameters specifically the filter dimensions (kernel size), the number of feature maps, and the stride length were globally optimized using the Bat Algorithm. The comprehensive simulation profiles verify the exceptional efficiency and classification performance of the presented framework.

Declaration

We acknowledge that we used ChatGPT to enhance the academic writing of our manuscript while ensuring the originality and integrity of our work.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Deep, S., & Zheng, X. (2019). Leveraging CNN and transfer learning for vision-based human activity recognition. In *2019 29th International Telecommunication Networks and Applications Conference (ITNAC)*. IEEE. https://doi.org/10.1109/ITNAC46935.2019.9078016

[2] Wang, C., & Peng, Z. (2020). Deep learning model for human activity recognition and prediction in smart homes. In *2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. IEEE. https://doi.org/10.1109/ICITBS49701.2020.00163

[3] Wang, S., & Zhu, X. (2020). A hybrid deep neural networks for sensor-based human activity recognition. In *2020 12th International Conference on Advanced Computational Intelligence (ICACI)*. IEEE. https://doi.org/10.1109/ICACI49185.2020.9177818

[4] Liu, Z., Xu, L., Jia, Y., & Guo, S. (2020). Human activity recognition based on deep learning with multi-spectrogram. In *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*. IEEE. https://doi.org/10.1109/ICSIP49896.2020.9339335

[5] Sarakon, S., & Tamee, K. (2020). An individual model for human activity recognition using transfer deep learning. In *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*. IEEE. https://doi.org/10.1109/ECTIDAMTNCON48261.2020.9090701

[6] Hristov, P., Manolova, A., & Boumbarov, O. (2020). Deep learning and SVM-based method for human activity recognition with skeleton data. In *2020 28th National Conference with International Participation (TELECOM)*. IEEE. https://doi.org/10.1109/TELECOM50385.2020.9299541

[7] Hossain Shuvo, M. M., Ahmed, N., Nouduri, K., & Palaniappan, K. (2020). A hybrid approach for human activity recognition with support vector machine and 1D convolutional neural network. In *2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE. https://doi.org/10.1109/AIPR50011.2020.9425332

[8] Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human activity recognition with inertial sensors using a deep learning approach. In *2016 IEEE Sensors*. IEEE. https://doi.org/10.1109/ICSENS.2016.7808590

[9] Ke, Y., Sukthankar, R., & Hebert, M. (2005). Efficient visual event detection using volumetric features. In *Tenth IEEE International Conference on Computer Vision (ICCV'05)* (Vol. 1, pp. 166–173). IEEE. https://doi.org/10.1109/ICCV.2005.85

[10] Schuldt, C., Laptev, I., & Caputo, B. (2004). Recognizing human actions: A local SVM approach. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)* (Vol. 3, pp. 32–35). IEEE. https://doi.org/10.1109/ICPR.2004.1334462

[11] Haque, A., Milstein, A., & Fei-Fei, L. (2020). Illuminating the dark spaces of healthcare with ambient intelligence. *Nature, 585*(7824), 193–202. https://doi.org/10.1038/s41586-020-2669-y

[12] Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters, 119*, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

[13] Zhu, R., Xiao, Z., Li, Y., Yang, M., Tan, Y., Zhou, L., ... Wen, H. (2019). Efficient human activity recognition solving the confusing activities via deep ensemble learning. *IEEE Access, 7*, 75490–75499. https://doi.org/10.1109/ACCESS.2019.2922104

[14] Xaviar, S., Yang, X., & Ardakanian, O. (2024). Centaur: Robust multimodal fusion for human activity recognition. *IEEE Sensors Journal, 24*(11), 18578–18591. https://doi.org/10.1109/JSEN.2024.3388893

[15] Dhekane, S. G., & Ploetz, T. (2025). Transfer learning in sensor-based human activity recognition: A survey. *ACM Computing Surveys, 57*(8), Article 1, 1–39. https://doi.org/10.1145/3717608

[16] Huang, J., & Ma, Y. (2020). Bat algorithm based on an integration strategy and Gaussian distribution. *Mathematical Problems in Engineering, 2020*, Article 9495281. https://doi.org/10.1155/2020/9495281

[17] Shehab, M., et al. (2022). A comprehensive review of bat inspired algorithm: Variants, applications, and hybridization. *Archives of Computational Methods in Engineering, 30*(2), 765–797. https://doi.org/10.1007/s11831-022-09817-5