An Adaptive Human Activity Recognition Framework Using CNN Architecture Design Schemes: A Survey

Daraksha Parveen[1], Sandeep Kumar Singh[2]

[1]*M. Tech Scholar, Dept. of CSE, Saraswati Higher Education & Technical College of Engineering, (AKTU), Varanasi, India*

[2]*Assistant Professor, Dept. of CSE, Saraswati Higher Education & Technical College of Engineering, (AKTU), Varanasi, India*

Abstract- Human Activity Recognition is an emerging field of research that focuses on identifying and interpreting physical activities performed by individuals through data collected from various sensors. With advancements in wearable devices, smartphones, and ambient sensors, HAR has gained significant importance in areas such as healthcare monitoring, smart homes, fitness tracking, elderly care, and human-computer interaction. The core of HAR lies in the effective collection, preprocessing, feature extraction, and classification of sensor data to accurately detect activities such as walking, running, sitting, standing, or more complex behaviors. Modern HAR systems employ machine learning and deep learning techniques to enhance recognition accuracy and robustness in dynamic environments. Despite considerable progress, HAR still faces challenges including inter-person variability, sensor placement issues, noise in data, and real-time processing demands. This research aims to explore various approaches and algorithms in HAR, evaluate their effectiveness, and propose improvements to enhance the accuracy and applicability of human activity detection systems in real-world scenarios.

Keywords - Human activity Recognition, Machine learning, Deep learning.

I. INTRODUCTION

In recent years, most nations have faced major issues related to the ageing population. It is challenging for their families and governments to care for these elderly individuals because some of them are forced to live alone. This is especially true when an emergency arises, for example falling. Early detection of these harmful practices is crucial to preventing those potential dangers or further damages [1]. Medical assistance can be sent out immediately if such incidents can be identified or even predicted. The goal of the research area known as "human activity recognition" (HAR) is to define and implement new methods for automatically identifying human activities using signals captured by wearable and/or environmental sensors. Most of the time, environmental devices must be installed in the home conditions, and equipment like cameras are seen as intrusive, specifically by elderly individuals. Consequently, the use of wearable devices has been increasingly focused in recent years. Among them cellphones, smartwatches and fitness equipment are currently receiving special attention. This is mostly because they are widely used by the public and gadgets have a variety of sensors built in (e.g., accelerometer, gyroscope, orientation and GPS). There are numerous applications for HAR approaches based on signals from the sensors of wearable devices [2]. Some of these fields include sport monitoring using accelerometer and GPS signals to assess user activity, fall detection using wearable technology to reduce elder mortality, behavioral analysis using physical measurements to avoid dementia disorders, among others. Being an important subfield of the computer vision, human motion recognition (HAR) has a lot of potential applications for creating products that enhance the level of comfort. It makes it possible for computers to interpret how a human behaves in a situation and, as a result, to take initiative in varying situations. Machine learning algorithms and the sensors used by each system can be used to classify human motion detection systems. Human action recognition can be viewed as a machine learning challenge. HAR systems can avoid this issue by extracting features from sensor data, building models for each action, and using these models to classify subsequent actions. Many supervised and unsupervised machine learning techniques have been employed over time for HAR [3]. Figure 1 shows a basic HAR architecture. The HAR algorithm has four parts that make up its overall workflow: acquiring sensor data, pre-processing it, extracting offline features, training a model, and finally classifying online activity.

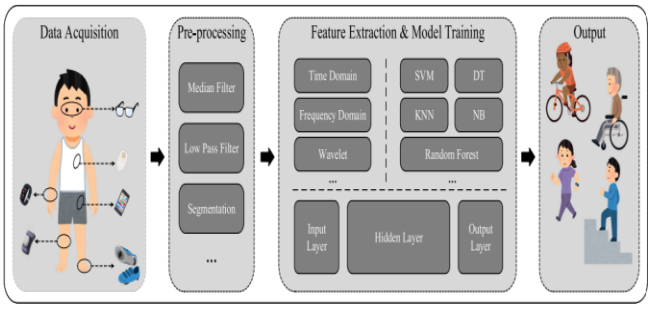

Figure 1: The processing flow of the human activity recognition system

The data acquisition phase involves the collection of data from Inertial Measurement Unit (IMU) sensors embedded in glasses, phones, watches or wrist bands, chest patches, shoes, etc. The collected data aptly describes the subject's actions in relation to certain bodily parts. The data that the sensor network collects about human movements is in constant flux. In terms of duration, there are two types of activities. The first kind comprises the activities which can be stationary or dynamic. The second generation is considered by brief activities, like postural adjustments. Median filters and low-pass filters are typical techniques for cleaning the data in the pre-processing step to remove noise interference and redundant information since the collected signal is prone to natural sensor drift and unconscious motions of the subject [4]. Moreover, continuous data segmentation is required at this stage, which involves splitting the signal into sliding windows with or without overlaps. The stage of feature extraction and model training is crucial for identifying important low-dimension patterns in the initial high-dimension sensor data. Current HAR solutions can be grouped into two categories based on distinct feature extraction techniques: hand crafted feature extraction with conventional machine learning and deep learning employing unsupervised features. The system uses an end-to-end training method to automatically mine these properties. The major emphasize of term "activity classification" is on creating an association amid extracted features and a particular kind of action with the help of classifiers. Human Activity Recognition techniques use a ML approach for predicting the data. It serves as a simple way to compare two streams of various lengths or rates. It is essential to enhance various classifiers effective for recognizing the human activity. A supervised learning model called Support Vector Machine (SVM) is employed for classification and regression analysis. SVM was created by Vapnik and is based on statistical learning theory [5]. This theory lays out a set of guidelines that must be followed in order to create classifiers with strong generalization abilities. The term "Support Vector Machine" highlights the significance of the closest vectors to the separation margin in the SVM's complexity. SVM was first developed to address binary problems, although there are also SVM formulations for multi-class issues. Nevertheless, due to its complexity, one-versus-one, one-versus-all, DAGSVM, and error-correcting code are the methods that are most frequently employed. The non-linear versions are produced using the kernel trick following the generation of linear SVM. A fundamental statistical strategy for solving the issue of pattern classification is known as Bayesian decision theory. This method is based on calculating the costs associated with different classification decisions and their associated probabilities. It assumes that the decision-making problem is framed in probabilistic terms and that all pertinent probabilities are known. RF is a hybrid method in which various DTs are constructed when the class is trained and output is received. This class illustrates the manner of the classes related to dissimilar trees [6]. K Nearest Neighbor (kNN) is a supervised learning technique. The fundamental principle of this approach is to classify the unknown sample by locating the k labelled samples that are closest to it. kNN training requires a minimal amount of work. On the other hand, labelling an unidentified sample has a rather significant computational cost. The parameter in kNN is the number of neighbours. Multi-layer neural networks, sometimes known as MLPs, are neural networks that contain one or more hidden layers. Due to the fact that it needs the required outcome to learn, this sort of neural network is known as a supervised learning network. The input layer, the hidden layer, and the output layer are the three sections of the MLP neural network. Layers of "input" units and "hidden" units are linked to layers of "output" units and layer of "hidden" units, respectively [7]. Every neuron in an ANN is connected to every other neuron in every layer and every neuron in the layer above it, making ANNs a completely connected network. A feed-forward network refers to a neural network architecture in which connections exist only between neurons in a given layer, and there are no connections that go back to neurons in previous layers. This network is known as a feed-forward network when the neurons are required to adjust the weights with the preceding layer. The autoencoder was first proposed as an unsupervised pre-training technique for artificial neural networks (ANN) in the 1980s under the name "auto associative learning module". A popular unsupervised technique for learning features is autoencoding. The results generated by autoencoders are frequently utilized as inputs for enhancing the performance of other networks and algorithms. Typically, an autoencoder is made up of two modules: an encoder and a decoder [8]. The initial unit of an autoencoder encodes input signals into a latent space, and the latter unit reconstructs signals from the latent space back into their original form. Many standard unsupervised networks are stacked to create a DBN, with each network's hidden layer acting as the visible layer for the layer following it. Basically, the restricted Boltzmann machine (RBM) represents each sub-network, which is an undirected generative energy-based framework featuring a "visible" input layer, a hidden layer, and intra-layer connections among them. In Deep Belief Networks (DBNs), there are typically connections that occur between different layers, although there are no connections between individual units within each layer. The design of DBNs enable a quick, unsupervised training procedure that can be done layer by layer. This is achieved by using contrastive divergence, a training strategy that estimates an association amid the network's weights and its error, which is implemented to each pair of layers sequentially from the lowest pair [9]. One of the earliest successes in efficient deep learning algorithms was the performance of DBNs in this method of training. A CNN is made up convolutional layers, pooling layers, fully connected layers, and an output layer. The convolutional layers are responsible for convolution operation. The process of learning space invariant features is made possible by the convolution function with a common kernel. In contrast to a fully-connected neural network, convolutional layers are better at capturing local dependency since each filter has a

specific receptive field (RF). Every kernel in a layer may only cover a small area of input neurons, however when several layers are combined [10], the neurons in higher layers are capable of jointly covering a larger area of the input, leading to a larger RF. The pyramidal design of a Convolutional Neural Network promotes its ability to aggregate the features of lower level into higher level. This algorithm is exploited for learning the classical attributes because of this property, by comparing the attributes derived from this algorithm to manual time and frequency domain attributes such as FFT and DCT. In majority of cases, CNN includes a pooling layer after every conv. layer.

II. LITERATURE REVIEW

M. M. Hossain Shuvo, et.al (2020) developed a hybrid adaptive framework for recognizing human activity [11]. A two-stage learning procedure was put forward for recognizing human activity, captured from a waist-mounted accelerometer and gyroscope sensor. Initially, the Random Forest (RF) binary algorithm was implemented for classifying the activity as inactive and dynamic. Subsequently, this framework focused on recognizing static activity using SVM, and adopted a one dimensional-CNN-based method to recognize dynamic activities. Hence, the developed framework performed more robustly and adaptively. Furthermore, the second method was proved effective for capturing local dependencies of activity signals and preserving the scale invariance. The results on UCI-HAR dataset, indicated that the developed framework offered an accuracy of 97.71% for recognizing 6 activities.

W. Shi, et.al (2022) investigated a multi-channel convolutional neural network with data augmentation (AMC-CNN) in order to recognize human activity [12]. First of all, the feature window was built using the sliding windows in time series. Moreover, the data was converted and inserted to augment the feature window. After that, a relatively lightweight MC-CNN algorithm was developed. The reasonable convolution kernel sets were deployed for mining the deep correlations among sensor data in a more effective way, so that MC convolutions were evaluated and parallel multi-scale features were extracted. This algorithm was trained for recognizing human activities. In the end, WISDM and MHEALTH datasets were applied to quantify the investigated algorithm in experimentation. The experimental results confirmed that the investigated algorithm offered efficiency for recognizing human activities based on single-sensor and multi-sensor at 99% accuracy.

S. E. Adi, et.al (2021) introduced a stacked Long Short-Term Memory (LSTM) model which recognized human actions on a smartphone [13]. An edge device was employed to process the data, exhibited that the transmission of raw data towards the cloud was not required to alleviate the potential bandwidth, energy consumption, and secrecy concerns. The offline prototype system offered an accuracy of 92.8% to classify six activities on a publicly available dataset. The quantization methods were useful for mitigating the weight representations. According to experiments, the introduced approach attained an accuracy of 92.7% and a memory footprint of 27 KB.

T. Ahmad, et.al (2023) suggested a method in which convolution neural network (CNN) and Bidirectional-gated recurrent unit (Bi-GRU) algorithms were employed for identifying human activity and proceeding the visual data [14]. The initial task was to extract the deep attributes from frames sequence of human activities videos based on CNN. The significant attributes were selected from the deep appearances for enhancing the efficiency and lessening the computing complexity. At second, Bi-GRU algorithm was put forward for learning the temporal motions of frames sequence, and the extracted attributes were inserted in this algorithm for learning the temporal dynamics at each time step. YouTube11, HMDB51 and UCF101 datasets were employed for computing the suggested method. The experimental outcomes validated the supremacy of the suggested method over the traditional methods.

R. K. Athota, et.al (2022) projected a couple of Hybrid Learning Algorithms (HLA) for developing the classifiers to recognize human activities on wearable sensor data [15]. CMFA and CGFA models were implemented for learning local attributes and long-term and gated-term dependencies in sequential data. Several filter sizes were utilized for improving the process of extracting features which were further exploited for capturing diverse local temporal dependencies. The WISDM dataset was executed to deploy the Amalgam Learning (AL) model. The projected algorithms offered an accuracy of 97.76% for smartwatch and 94.98% for smartphone using the initial model, and 96.91% for smartphone and 84.35% for smartphone using the second model. The experimental outcomes depicted that the projected algorithms outperformed the traditional methods.

A. Halim, et.al (2020) designed a framework in which RKELM model was exploited for recognizing human actions [16]. The embedding of an accelerometer sensor was done in a smartphone to compute the calories burned. Moreover, this framework was modified from ELM with the Gaussian kernel. Different metrics such as precision, recall, and F1score were considered for evaluating the designed framework on a dataset taken from London py data event in 2016. The simulation results indicated that the designed framework yielded an accuracy of 97.53%, F1 score up to 97% and consumed 0.06 secs below.

Y. Liu, et.al (2023) established a new device-free technique on the basis of Time-streaming Multiscale Transformer (TransTM) [17]. This technique supported the data fitting potentials of deploying un-processed Radio-Frequency Identification-Received Signal Strength Indicator (RFID-RSSI) data as input. In addition, the established technique

was assisted in capturing the behavioral attributes for recognizing the activities and interactions. In contrast to the traditional methods, this technique was scalable, generalized, and offered efficient data fitting. Thereafter, RF signals were utilized to provide higher efficacy for classifying human behavior. The datasets of real time were employed to evaluate the established technique. According to the experimental results, this technique attained an accuracy of 0.991.

K. K. Verma, et.al (2021) formulated the deep transfer learning (DTL) technique for extracting the attributes in conjunction with MC-SVM in order to classify human activities from colored videos [18]. A pre-trained CNN algorithm was suggested on the basis of VGGNet-19 for extracting the visual attributes from the RGB videos. Afterward, the extracted attributes were classified into diverse classes of human actions using MC-SVM. The UCF Sports Action dataset was employed to simulate the formulated technique. The simulation outcomes revealed that the formulated technique yielded an accuracy of 97.13%. Moreover, this technique was more effective in comparison with the existing methods.

K. Hirooka, et.al (2022) recommended a convolutional neural network (CNN) model for which an ensemble of transfer learning (TL) based multi-channel attention networks (MCAN) was implemented [19]. This model employed 4 CNN branches for generating feature fusion based ensembling. Besides, the contextual information was extracted from the feature map using an attention module in every branch. The last task of this model was to concatenate the extracted feature map. The fully connected network (FCN) exploited these maps for generating the final output to recognize human activity. The experiments were conducted on Stanford 40 actions, BU-101 and Willow datasets. The experimental results proved the superiority of the recommended model against the state-of-art methods.

2.1: Comparison Table

Author	Journal/Conference	Key Concepts	Accuracy	Future Enhancement
M. M. Hossain Shuvo, et.al (2020)	IEEE (AIPR)	developed a hybrid adaptive framework for recognizing human activity	Its accuracy was found 97.71%	The future plan would aim at deploying this framework into low-power integrated circuits for making it applicable on wearable sensors.
W. Shi, et.al (2022)	Journal (IEEE Access)	Investigated a multi-channel convolutional neural network with data augmentation (AMC-CNN) in order to recognize human activity	99% accuracy	The future work would aim to reduce the computing complexity and enhance the real-time interaction.
S. E. Adi, et.al (2021)	Conference (EMBC)	introduced a stacked Long Short-Term Memory (LSTM) model which recognized human actions on a smartphone	publicly available dataset-97.8% and 92.7% on other	This model would be enhanced to attain higher accuracy in real-time.
T. Ahmad, et.al (2023)	Journal (IEEE Access)	suggested a method in which convolution neural network (CNN) and Bidirectional-gated recurrent unit (Bi-GRU) algorithms were employed for	Accuracy was 93.8%	In future, this work would emphasize on recognizing human activity on internet of things (IOT) based devices

		identifying human activity		
R. K. Athota, et.al (2022)	Journal (Measurement: Sensors)	Projected a couple of Hybrid Learning Algorithms (HLA) for developing the classifiers to recognize human activities on wearable sensor data.	97.76% for smartwatch and 94.98% for smartphone using the initial model, and 96.91% for smartphone and 84.35% for smartphone using the second model.	This algorithm would be enhanced further to recognize the activities on large dataset.
A. Halim, et.al (2020)	Conference (ICIC)	designed a framework in which RKELM model was exploited for recognizing human actions	Accuracy of 97.53%	Future studies would concentrate on computing number of calories with regard to food, weight, and age for suggesting precise activities to users.
Y. Liu, et.al (2023)	Journal (Defence Technology)	established a new device-free technique on the basis of Time-streaming Multiscale Transformer (TransTM)	accuracy of 0.991	The issue of data imbalance would be resolved.
K. K. Verma, et.al (2021)	Conference (UPCON)	formulated the deep transfer learning (DTL) technique for extracting the attributes in conjunction with MC-SVM in order to classify human activities	accuracy of 97.13%	This algorithm would be improved to recognize all kinds of activities.
K. Hirooka, et.al (2022)	Journal (IEEE Access)	recommended a convolutional neural network (CNN) model for which an ensemble of transfer learning (TL) based multi-channel attention networks (MCAN) was implemented	Accuracy was 97.5%	The future objective was of attaining the skeletal joint points of human actions from still images for achieving a higher recognition rate

III. CONCLUSION

Human Activity Recognition (HAR) has become a pivotal technology in the development of intelligent and context-aware systems, enabling a wide range of applications from healthcare monitoring to smart living environments. With the integration of advanced sensors and the application of machine learning and deep learning models, HAR systems have achieved significant accuracy in detecting and classifying human behaviors. However, challenges such as sensor variability, activity similarity, noise in data, and generalization across different users still persist. Addressing these limitations through robust data preprocessing, model optimization, and the fusion of multimodal sensor data is essential for enhancing performance. As research in HAR continues to evolve, it holds the promise of creating more adaptive, personalized, and efficient solutions that can positively impact everyday life and contribute to the advancement of ubiquitous computing and human-centric technologies.

IV. REFERENCE

[1] Majdi Rawashdeh, Mohammed GH. Al Zamil, Ghulam Muhammad, “A knowledge-driven approach for activity recognition in smart homes based on activity profiling”, 2017, Future Generation Computer Systems

[2] Naoya Yoshimura, Takuya Maekawa, Takahiro Hara, “Preliminary Investigation of Visualizing Human Activity Recognition Neural Network”, 2019, Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU)

[3] Narjis Zehra, Syed Hamza Azeem, Muhammad Farhan, “Human Activity Recognition Through Ensemble Learning of Multiple Convolutional Neural Networks”, 2021, 55th Annual Conference on Information Sciences and Systems (CISS)

[4] Jiahui Huang, Shuisheng Lin, Ning Wang, Guanghai Dai, Yuxiang Xie, Jun Zhou, “TSE-CNN: A Two-Stage End-to-End CNN for Human Activity Recognition”, 2020, IEEE Journal of Biomedical and Health Informatics

[5] Nilay Tüfek, Ozen Özkaya, “A Comparative Research on Human Activity Recognition Using Deep Learning”, 2019, 27th Signal Processing and Communications Applications Conference (SIU)

[6] Hanyuan Xu, Zhibin Huang, Jue Wang, Zilu Kang, “Study on Fast Human Activity Recognition Based on Optimized Feature Selection”, 2017, 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES)

[7] Hairui Jia, Shuwei Chen, “Integrated data and knowledge driven methodology for human activity recognition”, 2020, Information Sciences

[8] Md Maruf Hossain Shuvo, Nafis Ahmed, Koundinya Nouduri, Kannappan Palaniappan, “A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network”, 2020, IEEE Applied Imagery Pattern Recognition Workshop (AIPR)

[9] Mohanad Babiker, Othman O. Khalifa, Kyaw Kyaw Htike, Aisha Hassan, Muhamed Zaharadeen, “Automated daily human activity recognition for video surveillance using neural network”, 2017, IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)

[10] Syed K. Bashar, Abdullah Al Fahim, Ki H. Chon, “Smartphone Based Human Activity Recognition with Feature Selection and Dense Neural Network”, 2020, 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)

[11] M. M. Hossain Shuvo, N. Ahmed, K. Nouduri and K. Palaniappan, "A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network," 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington DC, DC, USA, 2020, pp. 1-5

[12] W. Shi, X. Fang, G. Yang and J. Huang, "Human Activity Recognition Based on Multichannel Convolutional Neural Network with Data Augmentation," in IEEE Access, vol. 10, pp. 76596-76606, 2022

[13] S. E. Adi and A. J. Casson, "Design and optimization of a TensorFlow Lite deep learning neural network for human activity recognition on a smartphone," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 7028-7031

[14] T. Ahmad, J. Wu, H. S. Alwageed, F. Khan, J. Khan and Y. Lee, "Human Activity Recognition based on Deep-Temporal Learning using Convolution Neural Networks Features and Bidirectional Gated Recurrent Unit with Features Selection," in IEEE Access, vol. 34, no. 7, pp. 4335-4344, 2023

[15] R. K. Athota and D. Sumathi, “Human activity recognition based on hybrid learning algorithm for wearable sensor data”, Measurement: Sensors, vol. 10, no. 4, pp. 133-141, 14 October 2022

[16] A. Halim, E. Kwantan, S. Langie, V. Chandra and H. Gohzali, "Human Activity Recognition using Reduced Kernel Extreme Learning Machine for Body Weight Management," 2020 Fifth International Conference on Informatics and Computing (ICIC), Gorontalo, Indonesia, 2020, pp. 1-5

[17] Y. Liu, W. Huang and Y. Zhang, “TransTM: A device-free method based on time-streaming multiscale transformer for human activity recognition”, Defence Technology, vol. 7, no. 4, pp. 170003-170011, 23 February 2023

[18] K. K. Verma and B. Mohan Singh, "Vision based Human Activity Recognition using Deep Transfer Learning and Support Vector Machine," 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Dehradun, India, 2021, pp. 1-9

[19] K. Hirooka, M. A. M. Hasan, J. Shin and A. Y. Srizon, "Ensembled Transfer Learning Based Multichannel Attention Networks for Human Activity Recognition in Still Images," in IEEE Access, vol. 10, pp. 47051-47062, 2022