International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 243-248
Published Online May 2021 in IJEAST (http://www.ijeast.com)

AN ADVANCED APPROACH TO RECOGNIZE HUMAN ACTIVITIES VIA DEEP LEARNING

Aryan Karn, Dharm Raj Maurya
Motilal Nehru National Institute of Technology Allahabad, Prayagraj
Department of Electronics and Communication Engineering

Abstract- **The study of wearable and handheld sensors recognizing human activity improved our understanding of human behaviours and human objectives. Many academics seek to identify the activities of a user from raw data using the fewest necessary resources. In this article, we propose a network of profound beliefs, a full-service architecture for the recognition of activities (DBN-LSTM). This DBN-LSTM method improves the human predictability of raw data and reduces the complexity of the model as well as the requirement for comprehensive workmanship. A geographically and temporally rich network is CNN-LSTM. Our proposed model for the UCI HAR Public Data Set can achieve 99% accuracy and 92% precision.**

Keywords— **Computer vision, Artificial intelligence, Neural networks, CNN., Deep learning, Machine learning, Human Activity Detection**

I. INTRODUCTION

Due to their numerous practical applications in security, monitoring, and human-computer interaction, human behaviours in videos is also well studied. There are still significant variations in the aspects of activity and object in the dynamic contexts, together with a lack of annotated information, inaccurate definitions, and drifting of ideas. As not all training cases can be labelled and made available in advance, in the event of a surveillance or streaming challenge, you may need to gradually learn the models for your activity. In addition, there can always be new activities that provide vital data for improving models of activity. Current approaches to the recognition of human activity [1] are not in these cases because they are based on the presumption that all training sessions are predisposed and labelled. Strategies should therefore be developed to recognize online activities that work with streaming movies and newcomers. In addition, most contemporary techniques are based on models of handcrafted and static properties. The best for each application cannot be manually selected and static models and every application needs to be created individually. In addition, these model models cannot cope with changes in dynamic situations due to the static character of the feature model. One of the objectives of this project is to train unlabelled data modelling to recognize unregulated Internet activity. Due to its established theory and its high performance, since the introduction of deep learning, it has gained considerable interest in a number of computer vision applications [2]. Deep learning using technologies like turbot and autoencoder and stacking is used for the acquisition of meaningful, unattended, and supervised hierarchical characteristics [3] which are mostly hand-made features such as SIFT [4], HOG [5,], among others. We raise an essential question in this paper in the context of this discussion: Can one method be used indefinitely for profound learning activities models by streaming videos? Over time, new activities will appear and part of them will be marked by the active student. As a consequence, the total number of instances noted will increase. One naive method is to collect all these examples and use them from the beginning to train models of function and activity. But this strategy is unfeasible in a resource-constrained system due to a lack of storage capacity and computing power. We propose to use the supervised K-Medoid clustering approach in this study to identify the most informative subset of training cases. This allows us while maintaining the same performance as a system that uses all training instances, to save space and time.

II. RELATED WORK

In the literature, many HAR approaches were proposed. A different technique is applied for each of the two processes above. We divided the section in two parts in this connection: First and foremost,

R. Mutegeki et al [1] We compared their results with those of other methods as well. It competes with different designs and machine learning models of the deep neural network (DNN) that rely on previously suggested manual function data sets.

S. P. Pattar, et al [2] This module has a facial recognition system and a dialogue manager to enable personalized engagement. In relation to supervised learning methods, we discuss the benefits of autoencoders and how our proposed Architecture can be used in unscripted configurations to extend the duration of robot engagement. Experiments are also carried out with a humanoid robot in real-world interactions.

F. Wang, et al [3] This will lead to our research into state-of-the-art, deep learning technology. We show the inclusion of the LSTM network with channel selection to accurately identify the activity by blending the richer time and frequency features. The Intel 5300 Wireless Network Cards were used to build a prototype CSAR.

C. N. Phyo et al [4] This paper offers an intelligent HAR system that uses human skeleton information, imaging techniques, and

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 243-248
Published Online May 2021 in IJEAST (http://www.ijeast.com)

deeper learning to automatically recognize ordinary human actions from the deep sensor. Moreover, due to the low cost of calculation and good results of precision, the skeleton-based approach has proven extremely promising and can be used in any environmental or domain structural circumstances.

Y. Du, et al [5] The study describes a strategy to predict human conduct based on a profound learning model and evaluates how well it works with real data. Our approach is more predictive in comparison to the old technique. We will attempt to improve the accuracy of the prediction and add more activities in the future.

Zhang, Jin et al [6] One person may not be able to predict another person's actions by the model of activity recognition trained in. As only a small number of participants are registered on a given scale, recent studies have used a large amount of data from activity to train the model of recognition.

III. PROPOSED METHODOLOGY

The technique proposed is an improved accuracy with a new HAR DBN algorithm for the classification of human activity. For the beginning, frames are divided into human activity data sets in video sequences. The results are then turned into binary frames and morphological filtering is carried out to increase their quality. The new frames are then converted to a binary vector, which leads to an input matrix consisting of both the training and test data and their labels. As shown in Figure 1 this matrix shows data for our DBN architecture entrance. Finally, we train the DBN classifier to achieve the classification process by using the training data matrix. Due to the lighting on the background against the object in the frame, we used two approaches for the binarization stage. When the object was lighter than its background or vice versa, the threshold algorithm was used. On the other hand, it was used for frames with equivalent light degrees for the subject and background. the background detection approach. We then used morphological filters such as erosion, dilation, and more to remove the noise of the binary frame and finalize the image. Each binary frame returns a binary vector with the column counts equal to the column and line count product (i.e. the original frame size). A binary frame is provided as each binary vector takes a line in the matrix. Method DL-based solutions to the HAR problem were developed in recent years by a number of researchers. Extensive investigations are available. Several HAR researchers received attention from CNNs among the DL models. Smartphones are classified with a CNN dimension for the data from sensor activity. You compare your model with the performance of low-strength ML models such as SVM and DT. The results show that the model of CNN is more accurate. In category six daily seminars, CNN has employed a two-dimensional approach with a total of 12 volunteers. In terms of precision and overhead computation, your technology is compared to established machine learning algorithms. According to the results, both behaviours have improved. The new CNN method for classifying activities using two-dimensional CNNs is Ha et al. [19]. The performance of two CNN variations are compared, mainly based on the weights of the shared convolution layer. Another application for HAR CNN [11] has been submitted. The aim of the author is to examine different sensor setups for lower limbs and to identify the best placement of the sensor. [36] Authors use several RRN variants for the detection of abnormal behaviours, including GRUs and LSTMs, in order to detect daily activities in dementia aged. These models are comparable to flawless ML models in their performance. Most analysed measures (e.g. precision, precision, and reminder) show that RNNs are more effective than other ML models, with LSTMs showing somewhat better performance among the researched RNN models. This system's primary objective is a fair balance between energy consumption and accuracy of categorization. To do this, the system uses LSTMs to learn and magnetize movement to efficiently detect movement. As a field of study, deep education recently received considerable attention. The aim of this study is to find out how uncontrolled learning can be used to more abstractly represent input data. For a number of purposes, data, such as categorization and regression, can be used. The aim was to learn these profound representations using ordinary neural networks. However, it is not very easy to use descent gradients to train deep neural networks (i.e. multi-layered neural networks) [17]. The DBN has been able to tackle these problems by adding a greedy layer to an unregulated pre-training phase. This uncontrolled pre-workout creates an illustration by using downward gradient training to complete successful supervised learning [5, 17]. When a monitored descent is used, the network weights are shown in an unchecked step, which avoids local minima, differently from the random initialization. Hinton was the first to develop generative graphics models with DBN as an ANN architecture [1]. The DBN is a deep neural network with several layered variables both stochastic and latent, but no units in either layer (see Figure 2). The DBN model allows for probability, multilingual and gaussian data generation. The first of the two components of the DBN is composed of layers of reconstruction which turn data from input into abstraction. The second section consists of layers that convert this abstract image into classification labels for class prediction purposes.

IV. PROPOSED ALGORITHM

Restricted Boltzmann machines are particular varieties of Markov random fields. It is a Boltzmann machine that consists of an asymmetrical binary random unit network. Among the visible components of the network are:

1.Data are assigned RBM's w1 parameters in the first layer.

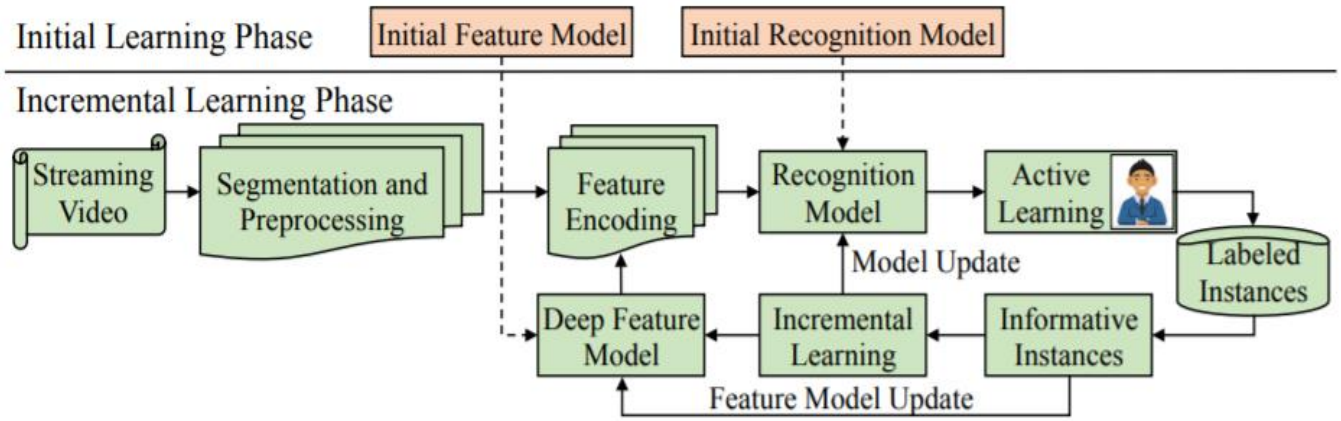
2.Setting up w1 and training the next layer of binary features in the RBM with the Q (h1 |v) = P(h1 |v; W1) samples.

3.Fixing test samples of h2 from Q(h2;h1) = P (h2;h1;W2) for w2, the second layer of teachers, and the teaching of the third layer.

4.Continuing this process recursively for the next layers.

training examples of class C are measured by nc
class C agents are identified by the number kc.
for each class, c. do
if kc <nc then
Randomly select kc data points {xk} of class
for k = 1: kc do
Compute {x
(k)

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 243-248
Published Online May 2021 in IJEAST (http://www.ijeast.com)



i
}, all the points nearest to xk.

Figure 1: Proposed Algorithm Model

Compute the median of the points {x
(k)
i
}.
end for
until no changes of xk.
else
Select all of the nc instances.
end if
end for

According to Hinton [4,], the DBN model is made up of one or more basic modules, each with its own RBM. In real-time behavioural modelling, input information is represented by a visible layer and hidden data features are displayed as a hidden layer. The result of this is also a speedy training technique by layer with their contradictory differences applied in the smallest layers (the smallest visible layer is the training set). Layer by layer, DBNs determines how variables are dependent on higher-level variables based on learning ascendant, generative weights. Essentially, DBNs involve deciding on the past distribution of visible vectors, p (v, h, W), as well as the past distribution of P (h, W) by the DBN weighing (W), which is determined by the RBM result. The probabilities of generating visible vectors are summarized as follows:

The nonlinear size of the system can be reduced by using them (v|h, W) is maintained, and p(h|W) is replaced by a superior rear vector combination model when W is found to be the weighted symmetry matrix between visible and hidden levels (i.e., the non-factorial distribution generated by averaging post-factorial distributions of each data vector). The hidden vectors of the training data are utilized as training data for a better model in the following training module. Every layer's latent value variable is created from the bottom up, starting with a data vector in the lower layer and generating weights in the opposite direction. There may be five major areas of completion for neural networks, making DBNs more multiple nonlinear layers are hidden within this Generative pre-training had been performed. Generativity pre-trained, they can be used to decrease the non-linear size when dealing with vector input. The networked teacher is another sensory input. There are two ways to use the DBN autoencoder and classifier. There are DBNs. We evaluated and used the DeeBNET Toolbox [18] for our proposed technique in MATLAB R2016b. A 64-bit Windows 10 PC was performed on all graphics cards: Intel Core i7-7700k CPU, NVIDIA GeForce GTX 980 (4 GB), and RAM 16GB (GDDR 4) .We used 18 video sequences from a range of training data settings for the KTH data set and 12 video sequences from various test data scenarios. We put DBMs architecture to the test in 10 human activities using the UIUC dataset (clapping, waving, running, jumping, walking, sitting-to-standing, turning, raising-one-hand, stretching-out). As a result, we have used 20 sequences of data images and 30 sequences of data in training.

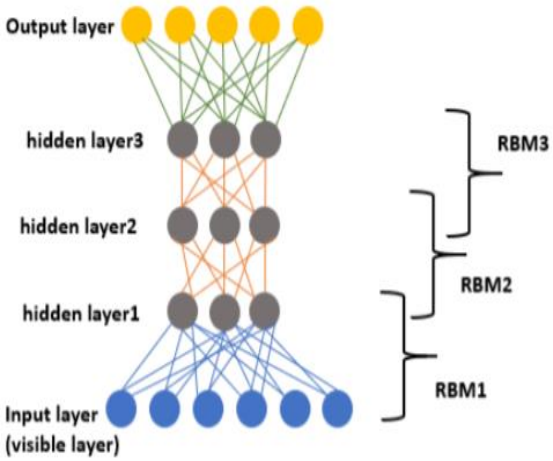


Figure 2: Neural Network Model

V. RESTRICTED BOLTZMANN MACHINE EQUATION

These are neural networks when they operate in an unsupervised or stochastic manner. According to Hinton [20], they can be used

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 243-248
Published Online May 2021 in IJEAST (http://www.ijeast.com)

to reduce size, classify, correlate, and conduct regressions. In each RBM, there are two levels, one that is visible and another that is hidden within the DBM. The two layers are connected and do not have connections within a layer. It is possible to determine the intrinsic relationship among binary data by using the properties of RBMs, whose energy functions are specified as follows:

$$p(v) = \sum_{h} p(h|W).p(v|h,W) \rightarrow (1)$$

$$E(h,v:o) = -\sum_{i=1}^{I}.\sum_{j=1}^{J}(w_{ij}v_{ij}h_j) - \sum_{i=1}^{I}(a_i v_i) - \sum_{i=1}^{I}(b_j h_j) \rightarrow (2)$$

$$p(h_j = 1|v;0) = \delta\left(\sum_{i=1}^{I} w_{ij}v_i + b_j\right) \rightarrow (3)$$

$$p(v_i = 1|h;0) = \delta\left(\sum_{j=1}^{J} w_{ij}h_j + a_i\right) \rightarrow (4)$$

Where, δ(x)=1/(1+exp(−x)). Gradient descent in the log-likelihood calculation can be performed by obtaining the following weights update:

$\triangle w_{ij} = \varepsilon\left((v_i h_j)data - (v_i h_j)model\right) \rightarrow (FinalEquation)$

VI. NEURAL NETWORK

These networks are made up of three neuronal units connected directly to their inputs. As a rule, there is a lesser number of hidden units than visible ones. Encoding (compression) and reconstructing (reconstruction) are the two steps in the auto-encoding process. We need to find the smallest error possible efficient way to represent the input data (i.e., a compact representation). DBN auto-encoders [18] are models containing auto-encoder regression-based models that are used to create generative models for extracting features from encrypted data. Data vectors are usually stored in the last hidden layer. Moreover, auto-encoders are a general class of algorithms used to reduce the size of input data representations.

In this study, a DBN-based classifier is used for teaching and supervised classification. This attention mechanism utilizes the input data feature vector of the very first layer of the feature map, the convolutional layers of the visible layer show the primitive detectors or reconstructions from the visible layer data, and the last layer of the DBN is the layer of the SoftMax, encompassing the classification labels. The first layer of the DBN is a visible layer of the layer. In order to use the classifier DBN architecture, it is extremely important to label the output data of the last RBM correctly. However, logistic regression can only be used for binary classifications. The DBN architecture plays an important role in a robust HAR system. Using this approach, our DBN-based HAR comprises a generative DBN, a generative RBM for the training phase, a generative practice RBM, and a discriminative practice RBM for classifying input data. In order to obtain the hidden layer's output, SoftMax regression is applied as the last layer. The DBN structure is shown in figure 3, the result of the layer-by-layer flow of raw data from the visible to the H3 hidden layers during DBN training three critical components of DBN-based HAR are created: training, fine-tuning, and classification. An initial coarse network is then the

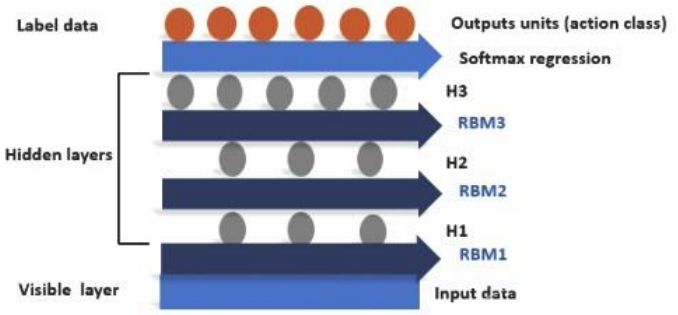


Figure 3: Network Model

initial coarse network is then constructed three critical components of DBN-based HAR are created: training, fine-tuning, and classification. An initial coarse network is then the initial coarse network is then constructed. By applying a contrastive version of Wake-Sleep, we are able to tailor the network weights using a top-down learning approach is used. We first input the normalized trajectory image, then classify it with SoftMax regression. An alternative method of multi-class classification uses generalized logistic regression. n only be used in binary classifications.

VII. DATASETS & RESULTS

Most HAR systems use the dataset [19] from KTH. There are six human actions in this video: boxing, handclapping, hand waving, jogging, running, and walking. 25 people perform the actions in four different scenarios (outside, outside with different scales, outside with different clothing, and inside). The KTH database, therefore, consists of 600 video segments shot with a static camera against homogeneous backgrounds.

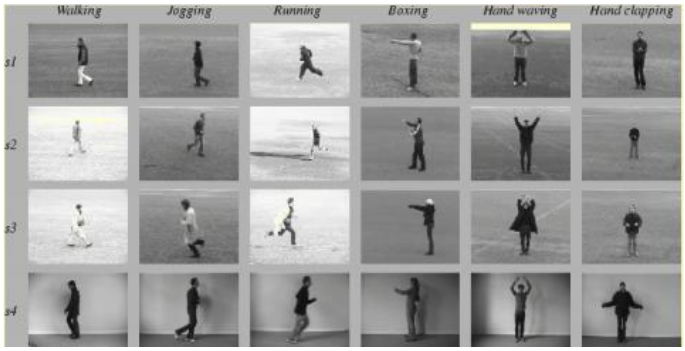


Figure 4: KTH Training Dataset

A 160*120-pixel resolution has been applied to each frame of each sequence stored in the AVI format. A sample image from the KTH dataset is shown in figure 4. Binary data is used for the DBN architecture in our proposed method. The majority of the

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 243-248
Published Online May 2021 in IJEAST (http://www.ijeast.com)

pre-processing involves converting the input data into binary data. By beginning with grayscale frames we segment each video sequence. Once the frames have been transformed into binary output, they are analysed. In order to enhance the quality of each frame, we apply a morphological filter. To ensure maximum compatibility, all frames are standardized to 95×55 pixels. By doing so, we can create an input matrix that consists of training and testing data, as well as labels for each frame.

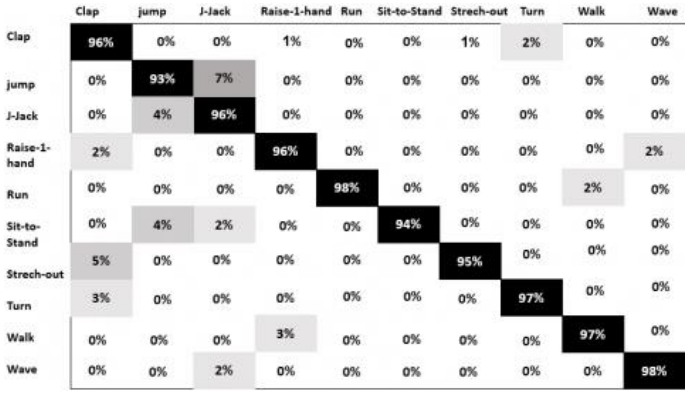


Figure 5: Confusion Matrix of **UCI HAR** DATASET

VIII. CONCLUSION AND FUTURE

This research proposes a new DensaLSTM profound learning model optimized for the smaller CSI data set and offers a baseline activity identification system that synthesizes diverse information on activities to reduce the effects and difficulties of activities. The motion of human limbs depends on the environment and time and varies in speed and scale. The human body has its features even for different people. The eight types of transformation methods employed by the system include drop-offs, gaussian sounds, time stretch, spectrum shifts, spectrum scaling, frequency filtering, and sample combinations. To prevent overruns and keep models compact, DenseLSTM combines and reuse's function mappings. HAR will incorporate data from motion capture systems and a DBN model for demonstrating time information in the future.

IX. REFERENCES

[1]. R. Mutegeki and D. S. Han, "A CNN-LSTM Approach to Human Activity Recognition," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 362-366, doi: 10.1109/ICAIIC48513.2020.9065078.

[2]. S. P. Pattar, E. Coronado, L. R. Ardila and G. Venture, "Intention and Engagement Recognition for Personalized Human-Robot Interaction, an integrated and Deep Learning approach," 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM), 2019, pp. 93-98, doi: 10.1109/ICARM.2019.8834226.

[3.]. F. Wang, W. Gong, J. Liu and K. Wu, "Channel Selective Activity Recognition with WiFi: A Deep Learning Approach Exploring Wideband Information," in IEEE Transactions on Network Science and Engineering, vol. 7, no. 1, pp. 181-192, 1 Jan.-March 2020, doi: 10.1109/TNSE.2018.2825144.

[4]. C. N. Phyo, T. T. Zin and P. Tin, "Deep Learning for Recognizing Human Activities Using Motions of Skeletal Joints," in IEEE Transactions on Consumer Electronics, vol. 65, no. 2, pp. 243-252, May 2019, doi: 10.1109/TCE.2019.2908986.

[5]. Y. Du, Y. Lim and Y. Tan, "Activity Prediction using LSTM in Smart Home," 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), 2019, pp. 918-919, doi: 10.1109/GCCE46687.2019.9015492.

[6]. Zhang, Jin; Wu, Fuxiang; Wei, Bo; Zhang, Qieshi; Huang, Hui; Shah, Syed W.; Cheng, Jun (2020). Data Augmentation and Dense-LSTM for Human Activity Recognition using WiFi Signal. IEEE Internet of Things Journal, (), 1–1. doi:10.1109/JIOT.2020.3026732

[7]. A. K. M. Masum, E. H. Bahadur, A. Shan-A-Alahi, M. A. Uz Zaman Chowdhury, M. R. Uddin and A. Al Noman, "Human Activity Recognition Using Accelerometer, Gyroscope and Magnetometer Sensors: Deep Neural Network Approaches," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944512.

[8]. M. Ronald, A. Poulose and D. S. Han, "iSPLInception: An Inception-ResNet Deep Learning Architecture for Human Activity Recognition," in IEEE Access, vol. 9, pp. 68985-69001, 2021, doi: 10.1109/ACCESS.2021.3078184.

[9]. M. N. Haque, M. Tanjid Hasan Tonmoy, S. Mahmud, A. A. Ali, M. Asif Hossain Khan and M. Shoyaib, "GRU-based Attention Mechanism for Human Activity Recognition," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934659.

[10]. H. Damirchi, R. Khorrambakht and H. D. Taghirad, "ARC-Net: Activity Recognition Through Capsules," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 1382-1388, doi: 10.1109/ICMLA51294.2020.00215.

[11]. H. Yan, B. Hu, G. Chen and E. Zhengyuan, "Real-Time Continuous Human Rehabilitation Action Recognition using OpenPose and FCN," 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), 2020, pp. 239-242, doi: 10.1109/AEMCSE50948.2020.00058.

[12]. M. Gochoo, T. Tan, S. Huang, S. Liu and F. S. Alnajjar, "DCNN-based elderly activity recognition using binary sensors," 2017 International Conference on Electrical and

Computing Technologies and Applications (ICECTA), 2017, pp. 1-5, doi: 10.1109/ICECTA.2017.8252040.

[13]. X. Fan, F. Wang, F. Wang, W. Gong and J. Liu, "When RFID Meets Deep Learning: Exploring Cognitive Intelligence for Activity Identification," in IEEE Wireless Communications, vol. 26, no. 3, pp. 19-25, June 2019, doi: 10.1109/MWC.2019.1800405.

[14]. Zewei Ding, Pichao Wang, P. O. Ogunbona and Wanqing Li, "Investigation of different skeleton features for CNN-based 3D action recognition," 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2017, pp. 617-622, doi: 10.1109/ICMEW.2017.8026286.

[15]. D. Martinelli, J. Cerbaro, J. A. Fabro, A. S. de Oliveira and M. A. SimoesTeixeira, "Human-robot interface for remote control via IoT communication using deep learning techniques for motion recognition," 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE), 2020, pp. 1-6, doi: 10.1109/LARS/SBR/WRE51543.2020.9307016.

[16]. Q. Liu et al., "Spectrum Analysis of EEG Signals Using CNN to Model Patient's Consciousness Level Based on Anesthesiologists' Experience," in IEEE Access, vol. 7, pp. 53731-53742, 2019, doi: 10.1109/ACCESS.2019.2912273.

[17]. J. Monteiro, J. P. Aires, R. Granada, R. C. Barros and F. Meneguzzi, "Virtual guide dog: An application to support visually-impaired people through deep convolutional neural networks," 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp. 2267-2274, doi: 10.1109/IJCNN.2017.7966130.

[18]. J. Zheng, R. Ranjan, C. Chen, J. Chen, C. D. Castillo and R. Chellappa, "An Automatic System for Unconstrained Video-Based Face Recognition," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 3, pp. 194-209, July 2020, doi: 10.1109/TBIOM.2020.2973504.

[19] Ijjina, E.P., Chalavadi, K.M. (2016). Human action recognition using genetic algorithms and convolutional neural networks. Pattern Recognition, 59: 199-212. http://dx.doi.org/10.1016/j.patcog.2016.01.012.

[20] Hinton, G.E. (2007). Boltzmann machine. Scholarpedia 2(5): 1668. http://dx.doi.org/10.4249/scholarpedia.1668