

Fire Hawk Optimizer with Deep Learning Enabled Human Activity Recognition

Mohammed Alonazi¹ and Mrim M. Alnfai^{2,*}

¹Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, 16273, Saudi Arabia

²Department of Information Technology, College of Computers and Information Technology, Taif University, Taif P.O. Box 11099, Taif, 21944, Saudi Arabia

*Corresponding Author: Mrim M. Alnfai. Email: m.alnofee@tu.edu.sa

Received: 06 July 2022; Accepted: 27 August 2022

Abstract: Human-Computer Interaction (HCI) is a sub-area within computer science focused on the study of the communication between people (users) and computers and the evaluation, implementation, and design of user interfaces for computer systems. HCI has accomplished effective incorporation of the human factors and software engineering of computing systems through the methods and concepts of cognitive science. Usability is an aspect of HCI dedicated to guaranteeing that human-computer communication is, amongst other things, efficient, effective, and sustaining for the user. Simultaneously, Human activity recognition (HAR) aim is to identify actions from a sequence of observations on the activities of subjects and the environmental conditions. The vision-based HAR study is the basis of several applications involving health care, HCI, and video surveillance. This article develops a Fire Hawk Optimizer with Deep Learning Enabled Activity Recognition (FHODL-AR) on HCI driven usability. In the presented FHODL-AR technique, the input images are investigated for the identification of different human activities. For feature extraction, a modified SqueezeNet model is introduced by the inclusion of few bypass connections to the SqueezeNet among Fire modules. Besides, the FHO algorithm is utilized as a hyperparameter optimization algorithm, which in turn boosts the classification performance. To detect and categorize different kinds of activities, probabilistic neural network (PNN) classifier is applied. The experimental validation of the FHODL-AR technique is tested using benchmark datasets, and the outcomes reported the improvements of the FHODL-AR technique over other recent approaches.

Keywords: Activity recognition; fire hawks optimizer; deep learning; usability; human computer interaction

1 Introduction

Usability and Human-Computer Interaction (HCI) are main aspects of the system development processes for enhancing and improving system facilities and for satisfying necessities and needs of users [1]. HCI would support users, designers, and analysts in identifying the requirements of system from



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

graphics, text style, color, fonts, and layout whereas usability would verify if the mechanism was easy to use, efficient, easy to learn, utility, easy to evaluate, easy to remember, and safe practical visible and offer job satisfaction to users [2]. HCI becomes a sub-field in computer science that deals with the study of the communication among people computers and users and the implementation, design, and evaluation of user interfaces for computer systems that were receptive to the user's habits and needs. It was a multidisciplinary domain, that involves design, computer science, and behavioral sciences. The main goal of HCI was to make computer systems user-friendly and highly usable [3]. Users communicate with computer system via user interfaces, that has software and hardware which offers means of input, permitting users for manipulating the output and system, permitting the system for providing data to the end-user [4]. The evaluation, design, and application of interfaces was been main aim of HCI. It can be identified in HCI whose good interface model presupposes a better theory or method of HCI, and that's why a theory must depend in large part on a theory of human cognition for modelling the cognitive process of users communicating with computer systems [5]. Different fields of HCI are displayed in Fig. 1.

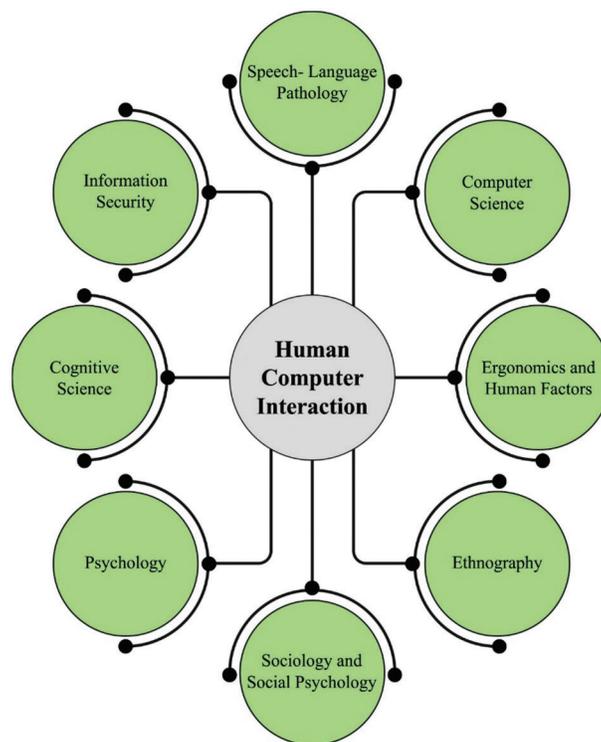


Figure 1: Different fields involved in HCI

Human activity recognition (HAR) serves an important role in human-to-human communication and inter-personal relationships. The reason behind offering information regarding the identity of an individual, their psychological state, and personality, it becomes tough for extraction [6]. The human capacity for recognizing another individual's actions becomes one chief subjects of research in the scientific regions of machine learning (ML) and computer vision (CV). So, in this research, many applications, which include robotics, video surveillance systems, and human-computer interaction for human behavior characterization, need multiple activity recognition mechanisms [7].

This article develops a Fire Hawk Optimizer with Deep Learning Enabled Activity Recognition (FHODL-AR) on HCI driven usability. In the presented FHODL-AR technique, the input images are

investigated for the identification of different human activities. For feature extraction, a modified SqueezeNet model is introduced by the inclusion of few bypass connections to the SqueezeNet among Fire modules. Besides, the FHO algorithm is utilized as a hyperparameter optimization algorithm, which in turn boosts the classification performance. To detect and categorize different kinds of activities, probabilistic neural network (PNN) classifier is applied. The experimental validation of the FHODL-AR technique is tested using benchmark datasets.

2 Related Works

Much prevailing research concentrates on feature extraction approaches due to the discriminative characteristics being significant to ensuring the generalizing ability of the HAR mechanism. There were 2 principal ways for extracting features from sensor-related data. One uses hand-crafted features related to the statistical knowledge; another one automatically derives features utilizing neural networks (NN) [8]. The derivation of meaningful hand-crafted features from time and frequency fields depends heavily on domain knowledge and human experience. Moreover, hand-crafted features were generally devised for a particular task and were not appropriate for more general tasks and atmospheres [9]. Deep learning (DL) advancements were broadly implemented due to DL methods could automatically extract high dimensional features and were not relies on field knowledge [10].

Ronald et al. [11] devised a convolutional neural architecture by ensembling transfer learning (TL) related multi-channel attention network. In this study, four convolutional neural networks (CNN) branches are utilized to make feature fusion related ensembling and in every branch, an attention method has been employed for extracting the contextual data from the feature map generated by prevailing pretrained methods. Lastly, the feature maps which is extracted from 4 branches concatenated, and put into fully-connected network for producing the final recognition output. Hirooka et al. [12] offer a new hybrid DL network for HAR which uses multi-modal sensor data, but this presented method was ConvLSTM pipeline which completely uses the data in every layer derived from the temporal domain. At last, a fully-connected layer and a softmax function were utilized for computing the probability of every class. Lv et al. [13] introduce a technique to recognize human actions utilizing skeleton data by RGB-D camera, called Kinect device. The HAR was learning in the CV field. In its application, the recognition of human activities is employed for image processing, sign language learning, surveillance of the elderly, and HCI. This technique depends upon skeleton data having coordinate value of every joint in human body, which is categorized utilizing SVM method when executing a movement for predicting the name of activities.

Komang et al. [14] modelled a new DL-related architecture for prediction and recognition of human activities related to a hybrid method. The foremost contribution of this work was to devise a novel hybrid architecture, compiling 4 wide-ranging pretrained network methods in an optimized way, utilizing a meta-heuristic technique. Yilmaz et al. [15] project a robust classifier method for HAR utilizing bidirectional long short-term memory (BiLSTM), wearable sensor data, and a hybrid of CNN can be leveraged. The devised multibranch CNN-BiLSTM network performs automated feature extraction from raw sensor data having minimum data preprocessing. The utility of BiLSTM and CNN forms the model to learn local features along with long-term dependencies in sequential data.

3 The Proposed Model

In this article, a new FHODL-AR technique has been developed for activity recognition on HCI driven usability. In the presented FHODL-AR technique, the input images are investigated for the identification of different human activities. The overall block diagram is shown in Fig. 2. Initially, the input data is

pre-processed and then the features are derived using the modified SqueezeNet model. Next, the recognition process take place using the PNN. Lastly, the FHO algorithm is used for parameter optimization.

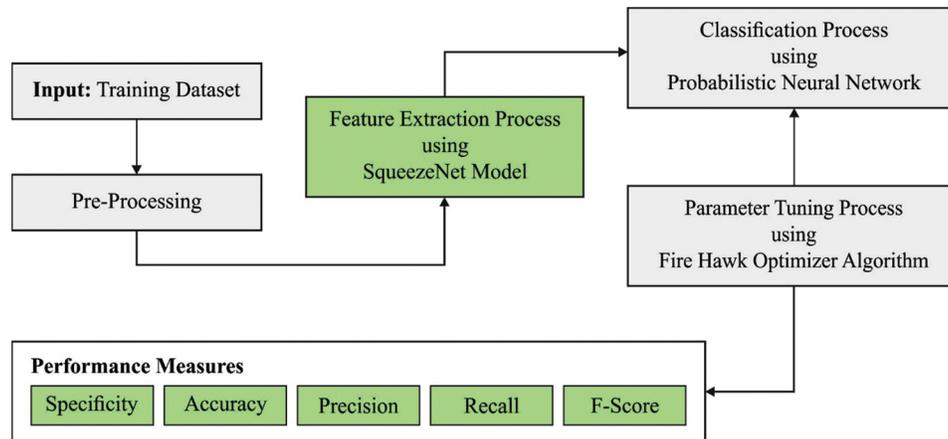


Figure 2: Block diagram of FHODL-AR model

3.1 Feature Extraction Using Modified SqueezeNet

For feature extraction, a modified SqueezeNet model is introduced by the inclusion of few bypass connections to the SqueezeNet among Fire modules [16]. The SqueezeNet is a small CNN structure that employs less parameter when preserving competitive performance. Various approaches are applied based on CNN to develop the SqueezeNet: (1) substitute 3×3 filters with 1×1 filter, (2) reduce the input channel number to 3×3 filters, and (3) down-sample in the network such that the convolutional layer has larger activation map. The SqueezeNet chiefly encompassed of Fire module that is squeeze convolutional layer with 1×1 filter. Then, these layer is fed into an extended layer that has a mixture of 1×1 and 3×3 convolutions. Here, a small CNN named SqueezeNet is applied that could accomplish more efficiency than other CNN architectures namely AlexNet when requiring less parameter that is useful in realtime scenarios. The fundamental SqueezeNet initiates by a convolutional layer and eight Fire modules that end with other convolutional layers. The filter count of each Fire module is progressively improved from the start to the ending of the network. The *max* pooling with a stride of 2 is implemented afterward layers conv1, Fire4, Fire8, and conv10. Rectified linear unit (ReLU) is adapted as the activation function and Dropout using a ratio of 0.5 is utilized afterward the Fire9 modules.

To increase the detection performance, an adapted SqueezeNet is developed by including bypass connection to the SqueezeNet among Fire modules. In this framework, bypass connection is additional nearby Fire modules 3, 5, 7, and 9, necessitating the module to learn a residual function among inputs and outputs. To perform a bypass connection around Fire3, we fixed the input to Fire4 equivalent to the output of Fire2 + output of Fire3, whereby the + operator is a component-wise calculation. These variations of the regularization employed to the parameter of the Fire module and, according to ResNet, might enhance the concluding accurateness or trainability.

3.2 Hyperparameter Tuning Using FHO Algorithm

Here, the FHO algorithm is utilized as a hyperparameter optimization algorithm, which in turn boosts the classification performance [17]. The FHO meta-heuristic approach stimulates the foraging behaviors of fire hawks that consider the procedure of catching prey, and setting and spreading fires. Firstly, a solution candidate (X) is defined by the location vector of the prey and fire hawks. An initial process is applied for identifying the first location of this vector in the searching region.

$$x = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}$$

$$\begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^j & \dots & x_1^d \\ x_2^1 & x_2^2 & \dots & x_2^j & \dots & x_2^d \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_i^1 & x_i^2 & \dots & x_i^j & \dots & x_i^d \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_N^1 & x_N^2 & \dots & x_N^j & \dots & x_N^d \end{bmatrix}, \begin{cases} i = 1, 2, \dots, N. \\ j = 1, 2, \dots, d. \end{cases} \tag{1}$$

$$x_i^j(0) = x_{i, \min}^j + \text{rand} \cdot (x_{i, \max}^j - x_{i, \min}^j), \begin{cases} i = 1, 2, \dots, N. \\ j = 1, 2, \dots, d. \end{cases} \tag{2}$$

Now, X_i denotes the i -th candidate solution in the searching region; d indicates the dimension vector; N denotes the overall amount of candidate solution in the searching region; x_i^j represent the j -th decision parameter of the i -th candidate solution; $x_i^j(0)$ characterizes the initialized location of the candidate solution; $x_{i, \min}^j$ and $x_{i, \max}^j$ represents the minimal and maximal limits of the j -th decision parameter for the i -th candidate solution; rand is a random number distributed uniformly within zero and one. Here, global optimal solution is considered as the major fire that is firstly applied by the Fire Hawks to b , the schematic representation of the aspect is given that is arithmetically given in the following:

$$PR = \begin{bmatrix} PR_1 \\ PR_2 \\ \vdots \\ PR_k \\ \vdots \\ PR_m \end{bmatrix} \quad k = 1, 2, \dots, m, \tag{3}$$

$$FH = \begin{bmatrix} FH_1 \\ FH_2 \\ \vdots \\ FH_l \\ \vdots \\ FH_n \end{bmatrix} \quad L = 1, 2, \dots, n, \tag{4}$$

Let PR_k be the k -th prey in the searching space w.r.t. overall amount of m preys; and FH_l denotes the l -th fire hawk considered an overall amount of n fire hawks in the searching region. Then, the overall distance between the prey and the Fire Hawks is evaluated.

$$D_k^l = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \begin{cases} l = 1, 2, \dots, n. \\ k = 1, 2, \dots, m. \end{cases} \quad (5)$$

In Eq. (5), D_k^l denotes the overall distance between l -th fire hawk and the k -th prey; m indicates the overall amount of prey in the searching domain; n represents the overall amount of fire hawks in the searching domain, and (x_1, y_1) and (x_2, y_2) characterize the coordinate of Fire Hawks and prey in the searching domain. In the meantime, birds are eager to apply the burning sticks from Fire Hawk territory; for that reason, those behaviors are applied as location update procedure in the major searching loop of FHO, as follows:

$$FH_1^{new} = FH_1 + (r_1 \times GB - r_2 \times FH_{Near}), \quad l = 1, 2, \dots, n, \quad (6)$$

In Eq. (6), FH_1^{new} indicates the novel location vector of l -th Fire Hawks (FH_1); GB denotes the global optimal solution in the searching domain assumed as the main fire; FH_{Near} denotes the other Fire Hawks in the searching domain; and r_1 and r_2 are random numbers distributed uniformly within $(0, 1)$ to define the movement of Fire Hawk towards the major fire and the other Fire Hawks' territory.

Then, the motion of prey inside the territory of Fire Hawk assumed a major aspect of animal activities for the location updating method.

$$PR_q^{new} = PR_q + (r_3 \times FH_1 - r_4 \times SP_1), \begin{cases} l = 1, 2, \dots, n. \\ q = 1, 2, \dots, r. \end{cases} \quad (7)$$

In Eq. (7), PR_q^{new} indicates the novel location vector of q th prey (PR_q) enclosed by l -th Fire Hawk (FH_1); GB shows the global optimal solution in the searching domain assumed as major fire; SP_1 indicate a safer position under l -th Fire Hawk territory; and r_3 and r_4 are random numbers distributed uniformly with $(0, 1)$ to define the motion of prey towards the Fire Hawk and the safer position.

In addition, the prey motion toward the other Fire Hawk territories while there is a chance the prey might be approaching the Fire Hawk in the nearby ambushes or hide in a safe position outside the Fire Hawk territories where they are trapped:

$$PR_q^{new} = PR_q + (r_5 \times FH_{Alter} - r_6 \times SP), \begin{cases} l = 1, 2, \dots, r. \\ q = 1, 2, \dots, n. \end{cases} \quad (8)$$

In Eq. (8), PR_q^{new} denotes the novel location vector of q th prey (PR_q) enclosed by the l -th fire hawk (FH_1); FH_{Alter} denotes the fire hawk in the searching domain; SP indicates a safer potion outside the l -th Fire Hawk territories; r_5 and r_6 are random numbers distributed uniformly within $(0, 1)$ to define the movement of prey towards the Fire Hawk and the safer location outside the territory and it is mathematically expressed in the following:

$$SP_1 = \frac{\sum_{q=1}^r PR_q}{r}, \begin{cases} q = 1, 2, \dots, r. \\ l = 1, 2, \dots, n. \end{cases} \quad (9)$$

$$SP = \frac{\sum_{k=1}^m PR_k}{m}, \quad k = 1, 2, \dots, m. \quad (10)$$

Now, PR_q indicate the q th prey bounded by the l -th fire hawk (FH_1); PR_k denotes the k -th prey in the searching domain.

3.3 Activity Recognition Using PNN

To detect and categorize different kinds of activities, PNN classifier is applied [18]. The PNN is an effective neural network that is commonly employed for object classification. Usually, PNN is more accurate and faster when compared to the multiple layer perceptron network, and is comparatively impervious to the outlier. PNN makes use of the Parzen estimator to estimate the probability distribution function (PDF) of all the classes. The multi-variate Bayesian rule is employed to distribute the class with the maximum posterior possibility to original input dataset. The architecture of PNN is discussed in the following. Given that $f_1(x)$ and $f_2(x)$ are PDF related to p-dimension input vector X for π_1 and π_2 , the previous probability and misclassification cost ratios are formulated below

$$X \in \left\{ \frac{\pi_1 f_1(x)/f_2(x)}{\pi_2} \geq \{ [C(1|2)/C(2|1)] \times [P_2/P_1] \} \right\} \quad (11)$$

In Eq. (5), $C(i|j)$ denotes the misclassification cost (single object is categorized into π_i however, it belongs to π_j). P_j denotes the previous possibility of existence of population π_j . The PDF is applied for estimating the posterior possibility that x belongs to class π_i . The PDF in PNN can be resolved using the Bayesian classification with Parzen estimator and it is formulated in the following equation:

$$f_A(X) = \frac{1}{(2\pi)^{p/2} b^p} \times \frac{1}{m} \times \sum_{i=1}^m \exp \left[-\frac{(X - X_{Ai})^T (X - X_{Ai})}{2b^2} \right] \quad (12)$$

In Eq. (6), i indicates the pattern count, m shows the overall amount of training patterns, X_{Ai} denotes i -th training pattern from π_A , and p denotes the dimension of input space. $f_A(X)$ be able to estimate smooth density function. There exist 4 layers in PNN: output, input, pattern, and summation layers.

The input unit is distribution unit that supplies a similar input value to the pattern unit and creates a dot product of pattern vector X using weight vector $W_i (Z_i = X * W_i)$. A non-linear function ($\exp [(Z_i - 1)/b^2]$) has been implemented on Z beforehand transmitting the activation level towards the summation layer. Consider that X and W_i are standardized towards a unit length, the non-linear function is formulated below

$$NO \sim \exp \left[-\frac{(X - W_i)^T (X - W_i)}{2b^2} \right] \quad (13)$$

Then, the non-linear function is a similar method to a Parzen estimator with a Gaussian kernel. The summation units sum the output for pattern unit respective to the class and evaluate the PDF. The output pattern applied the maximum vote to forecast the target class. Because the input layer is applied to the connection weight, PNN doesn't require altering the connection weight. As a result, the training speed is quicker when compared to the conventional backpropagation neural network (BP-NN).

4 Results and Discussion

This section inspects the HAR outcomes of the FHODL-AR model on two datasets. The first KTH dataset (<http://www.nada.kth.se/cvap/actions/>) comprises 600 samples under six class labels. Table 1 reports the details related to the KTH dataset. Next, the UCF Sports dataset comprises 1000 samples under ten classes (http://crcv.ucf.edu/data/UCF_Sports_Action.php). The details relevant to the UCF Sports dataset are given in Table 2.

The confusion matrices produced by the FHODL-AR model on KTH dataset are portrayed in Fig. 3. The results indicated that the FHODL-AR model has proficiently recognized all the class labels.

Table 1: Details of KTH dataset

Label	Class	No. of samples
C-1	Boxing	100
C-2	Handclapping	100
C-3	Handwaving	100
C-4	Jogging	100
C-5	Running	100
C-6	Walking	100
Total number of Samples		600

Table 2: Details of UCF sport dataset

Label	Class	No. of samples
C-1	Diving-Side	100
C-2	Golf-Swing	100
C-3	Kicking-Front	100
C-4	Lifting	100
C-5	Riding Horse	100
C-6	Run-Side	100
C-7	StateBoarding-Front	100
C-8	Swing-Bench	100
C-9	Swing-SideAngle	100
C-10	Walk-Front	100
Total Number of Samples		1000

Table 3 exhibits detailed HAR outcomes of the FHODL-AR model under distinct aspects of KTH dataset. The results implied that the FHODL-AR model has shown enhanced outcomes. With 80% of TR data, the FHODL-AR model has obtained average $accu_y$ of 99.31%, $prec_n$ of 98%, $reca_l$ of 97.90%, $spec_y$ of 99.58%, and F_{score} of 97.91%. Likewise, With 20% of TR dataset, the FHODL-AR method has attained average $accu_y$ of 99.44%, $prec_n$ of 98.52%, $reca_l$ of 98.45%, $spec_y$ of 99.66%, and F_{score} of 98.47%. Ultimately, With 70% of TR data, the FHODL-AR technique has achieved average $accu_y$ of 97.86%, $prec_n$ of 93.54%, $reca_l$ of 93.52%, $spec_y$ of 98.72%, and F_{score} of 93.47%. Meanwhile, With 30% of TR data, the FHODL-AR approach has gained average $accu_y$ of 99.07%, $prec_n$ of 97.11%, $reca_l$ of 97.10%, $spec_y$ of 99.45%, and F_{score} of 97.08%.

The training accuracy (TA) and validation accuracy (VA) achieved by the FHODL-AR algorithm on KTH dataset is displayed in **Fig. 4**. The experimental outcome denoted the DSOCDBN-STC system has established highest values of TA and VA. Especially, the VA is larger than TA.

The training loss (TL) and validation loss (VL) acquired by FHODL-AR process on KTH dataset are shown in **Fig. 5**. The experimental outcome implicit the FHODL-AR technique has attained minimum values of TL and VL. Especially, the VL is smaller than TL.

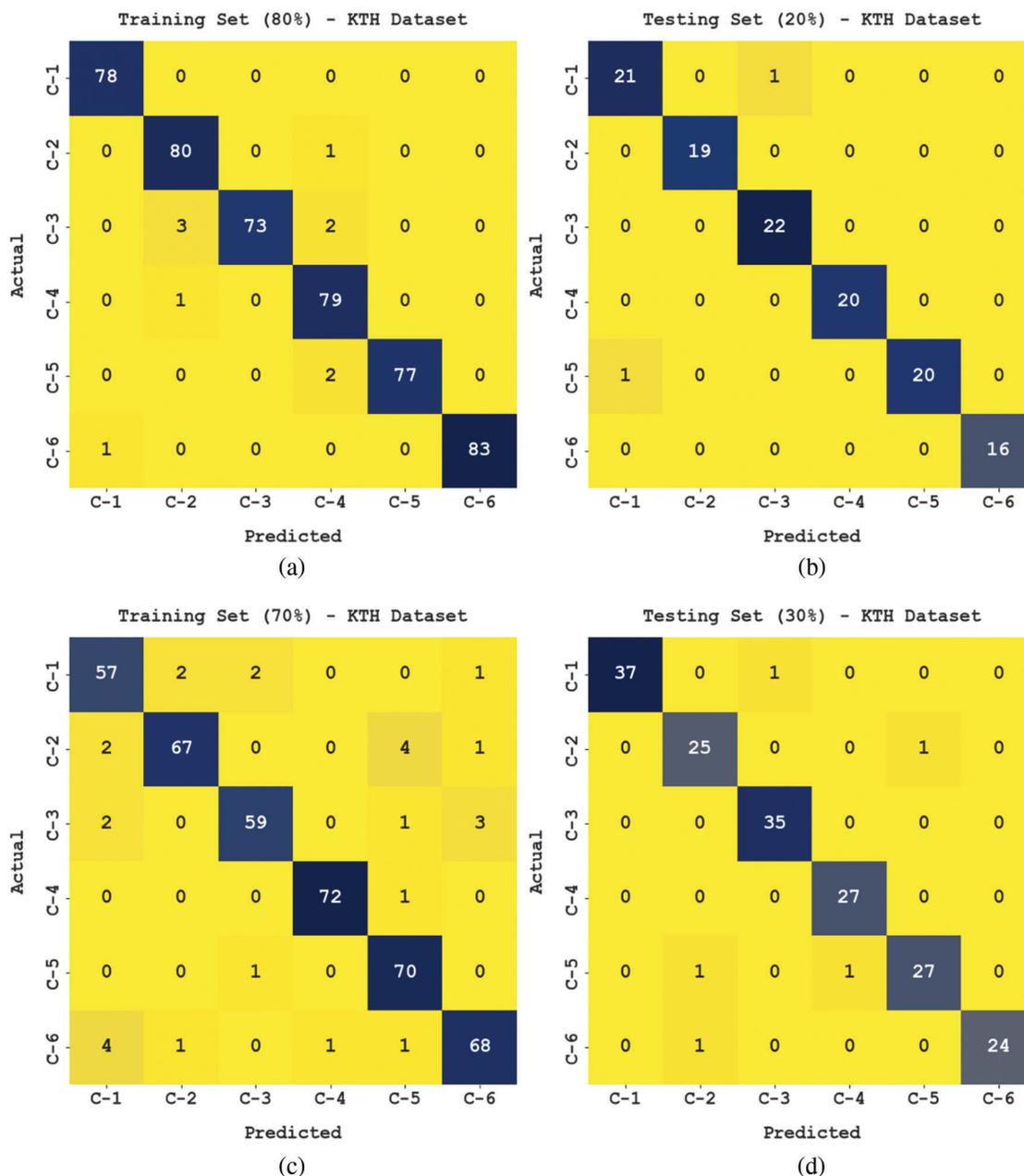


Figure 3: Confusion matrices of FHODL-AR model on KTH dataset

Table 4 and Fig. 6 exhibit a comprehensive $accu_y$ examination of the FHODL-AR model with other HAR models on KTH dataset. The experimental outcomes implied that the DBN and GRU models have shown poor performance with lower $accu_y$ values of 97.72% and 97.05% respectively. Next, the Bi-LSTM model has attained slightly enhanced $accu_y$ of 98.19%. Then, the RNN and CNN models demonstrated reasonable $accu_y$ of 98.62% and 98.60% respectively. However, the FHODL-AR model has accomplished superior $accu_y$ of 99.44%.

Table 3: Overall HAR outcomes of FHODL-AR model on KTH dataset

Labels	Accuracy	Precision	Recall	Specificity	F-score
Training set (80%)					
C-1	99.79	98.73	100.00	99.75	99.36
C-2	98.96	95.24	98.77	99.00	96.97
C-3	98.96	100.00	93.59	100.00	96.69
C-4	98.75	94.05	98.75	98.75	96.34
C-5	99.58	100.00	97.47	100.00	98.72
C-6	99.79	100.00	98.81	100.00	99.40
Average	99.31	98.00	97.90	99.58	97.91
Testing set (20%)					
C-1	98.33	95.45	95.45	98.98	95.45
C-2	100.00	100.00	100.00	100.00	100.00
C-3	99.17	95.65	100.00	98.98	97.78
C-4	100.00	100.00	100.00	100.00	100.00
C-5	99.17	100.00	95.24	100.00	97.56
C-6	100.00	100.00	100.00	100.00	100.00
Average	99.44	98.52	98.45	99.66	98.47
Training set (70%)					
C-1	96.90	87.69	91.94	97.77	89.76
C-2	97.62	95.71	90.54	99.13	93.06
C-3	97.86	95.16	90.77	99.15	92.91
C-4	99.52	98.63	98.63	99.71	98.63
C-5	98.10	90.91	98.59	97.99	94.59
C-6	97.14	93.15	90.67	98.55	91.89
Average	97.86	93.54	93.52	98.72	93.47
Testing set (30%)					
C-1	99.44	100.00	97.37	100.00	98.67
C-2	98.33	92.59	96.15	98.70	94.34
C-3	99.44	97.22	100.00	99.31	98.59
C-4	99.44	96.43	100.00	99.35	98.18
C-5	98.33	96.43	93.10	99.34	94.74
C-6	99.44	100.00	96.00	100.00	97.96
Average	99.07	97.11	97.10	99.45	97.08

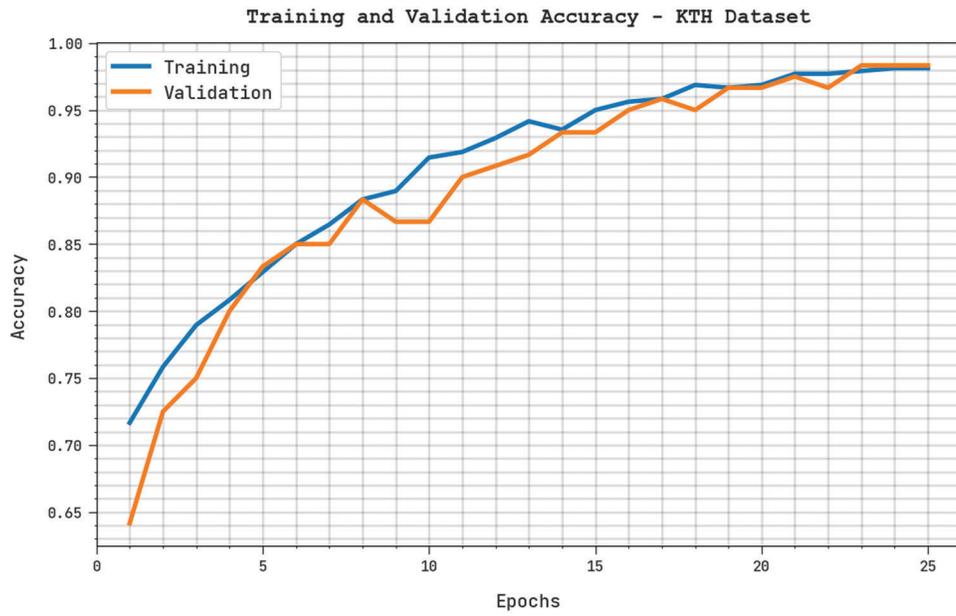


Figure 4: TA and VA of FHODL-AR model on KTH dataset

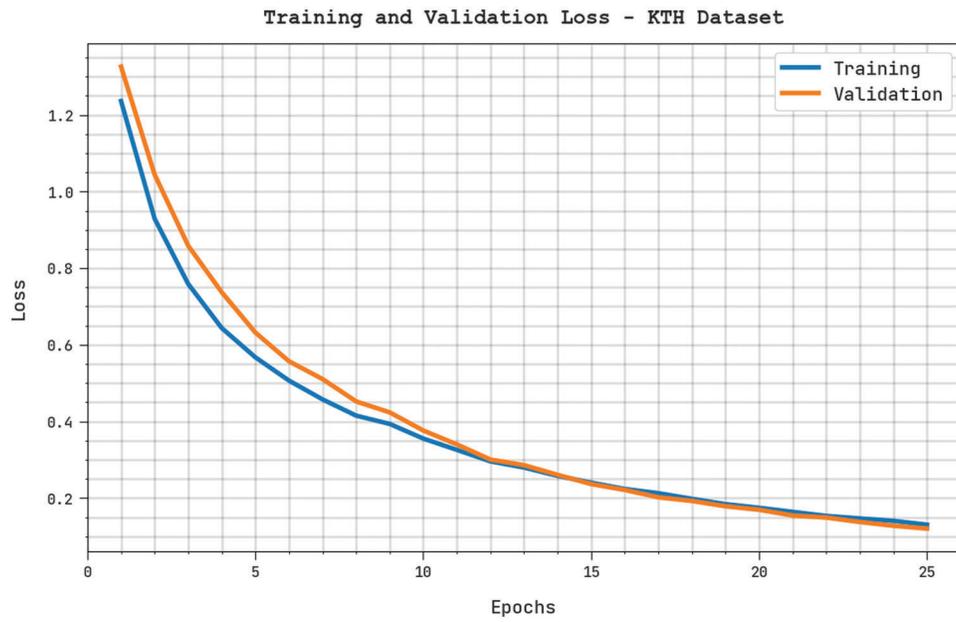
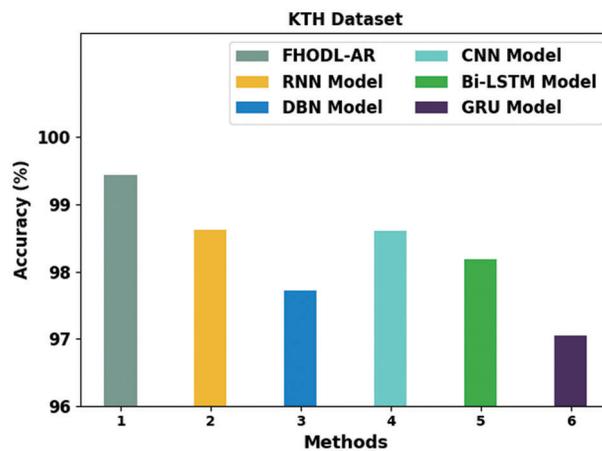


Figure 5: TL and VL of FHODL-AR model on KTH dataset

Table 4: Comparative $accu_y$ examination of FHODL-AR model on KTH dataset

Methods	Accuracy (%)
FHODL-AR	99.44
RNN model	98.62
DBN model	97.72
CNN model	98.60
Bi-LSTM model	98.19
GRU model	97.05

**Figure 6:** $Accu_y$ analysis of FHODL-AR model with existing models on KTH dataset

The confusion matrices produced by the FHODL-AR technique on UCF Sports data are indicated in Fig. 7. The outcomes portrayed that the FHODL-AR approach has proficiently identified each class label.

Table 5 displays comprehensive HAR outcomes of the FHODL-AR method under distinct aspects of UCF sports dataset. The results inferred that the FHODL-AR technique has shown better results. With 80% of TR dataset, the FHODL-AR approach has acquired average $accu_y$ of 99.22%, $prec_n$ of 96.10%, $reca_l$ of 96.10%, $spec_y$ of 99.57%, and F_{score} of 96.09%. Likewise, With 20% of TR dataset, the FHODL-AR algorithm has attained average $accu_y$ of 99.10%, $prec_n$ of 94.68%, $reca_l$ of 95.33%, $spec_y$ of 99.51%, and F_{score} of 94.94%. Eventually, With 70% of TR dataset, the FHODL-AR process has achieved average $accu_y$ of 98.63%, $prec_n$ of 93.31%, $reca_l$ of 93.11%, $spec_y$ of 93.24%, and F_{score} of 93.14%. Meanwhile, With 30% of TR dataset, the FHODL-AR system has acquired average $accu_y$ of 98.67%, $prec_n$ of 93.62%, $reca_l$ of 93.70%, $spec_y$ of 99.26%, and F_{score} of 93.56%.

Table 6 and Fig. 8 show a comprehensive $accu_y$ inspection of the FHODL-AR with other HAR techniques on UCF sports data. The experimental outcome implies that the Bi-LSTM and RNN approaches have shown poor performance with lesser $accu_y$ values of 97.01% and 97.60% correspondingly.

Then, the DBN technique has accomplished somewhat better $accu_y$ of 97.74%. Next, the GRU and CNN algorithms have established reasonable $accu_y$ of 98.21% and 98.26% correspondingly. However, the FHODL-AR system has obtained greater $accu_y$ of 99.10%. Therefore, the FHODL-AR method can be employed as a powerful mechanism for activity recognition.

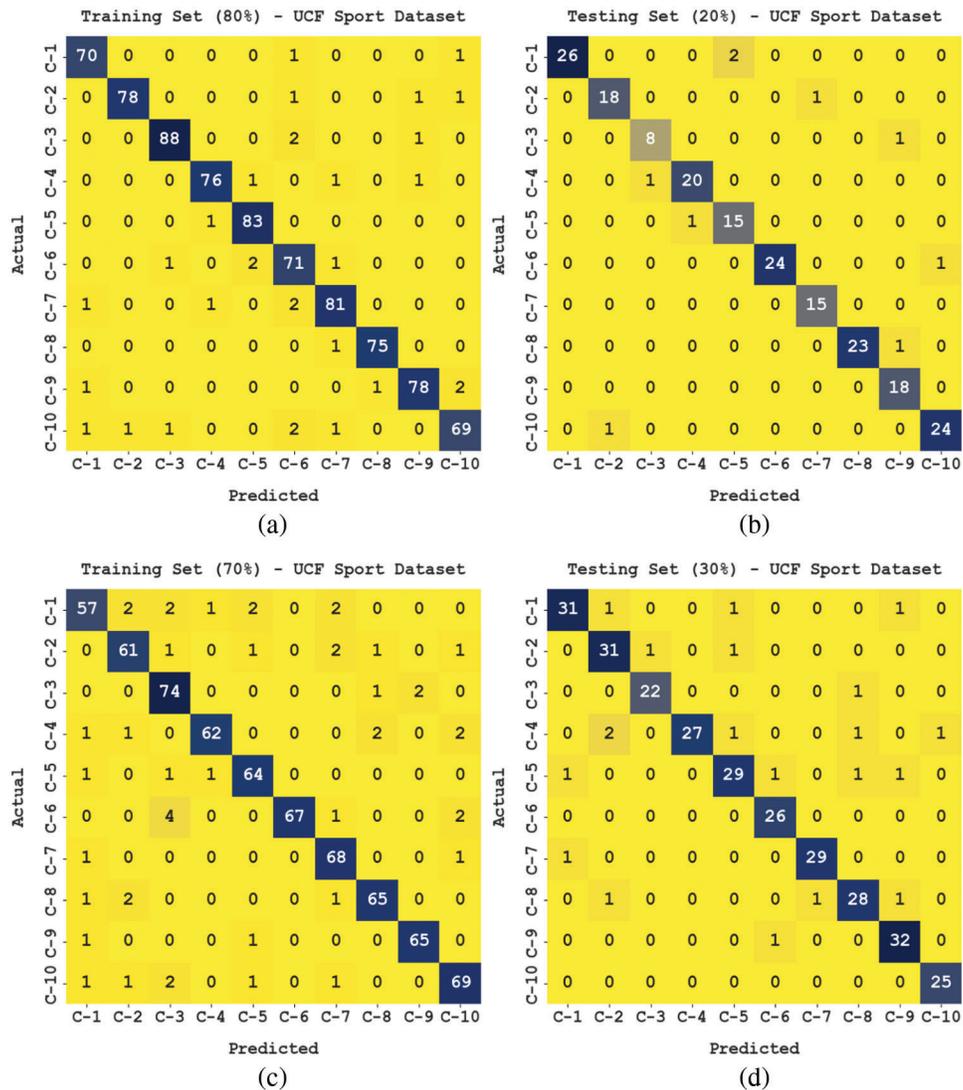


Figure 7: Confusion matrices of FHODL-AR model on UCF Sports dataset

Table 5: Overall HAR outcomes of FHODL-AR model on UCF Sport dataset

Labels	Accuracy	Precision	Recall	Specificity	F-score
Training set (80%)					
C-1	99.38	95.89	97.22	99.59	96.55
C-2	99.50	98.73	96.30	99.86	97.50
C-3	99.38	97.78	96.70	99.72	97.24
C-4	99.38	97.44	96.20	99.72	96.82
C-5	99.50	96.51	98.81	99.58	97.65
C-6	98.50	89.87	94.67	98.90	92.21
C-7	99.00	95.29	95.29	99.44	95.29
C-8	99.75	98.68	98.68	99.86	98.68

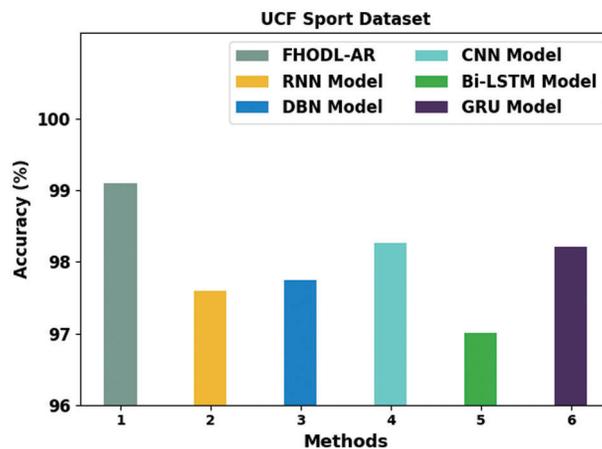
(Continued)

Table 5 (continued)

Labels	Accuracy	Precision	Recall	Specificity	F-score
C-9	99.12	96.30	95.12	99.58	95.71
C-10	98.75	94.52	92.00	99.45	93.24
Average	99.22	96.10	96.10	99.57	96.09
Testing set (20%)					
C-1	99.00	100.00	92.86	100.00	96.30
C-2	99.00	94.74	94.74	99.45	94.74
C-3	99.00	88.89	88.89	99.48	88.89
C-4	99.00	95.24	95.24	99.44	95.24
C-5	98.50	88.24	93.75	98.91	90.91
C-6	99.50	100.00	96.00	100.00	97.96
C-7	99.50	93.75	100.00	99.46	96.77
C-8	99.50	100.00	95.83	100.00	97.87
C-9	99.00	90.00	100.00	98.90	94.74
C-10	99.00	96.00	96.00	99.43	96.00
Average	99.10	94.68	95.33	99.51	94.94
Training set (70%)					
C-1	97.86	90.48	86.36	99.05	88.37
C-2	98.29	91.04	91.04	99.05	91.04
C-3	98.14	88.10	96.10	98.39	91.93
C-4	98.86	96.88	91.18	99.68	93.94
C-5	98.86	92.75	95.52	99.21	94.12
C-6	99.00	100.00	90.54	100.00	95.04
C-7	98.71	90.67	97.14	98.89	93.79
C-8	98.86	94.20	94.20	99.37	94.20
C-9	99.43	97.01	97.01	99.68	97.01
C-10	98.29	92.00	92.00	99.04	92.00
Average	98.63	93.31	93.11	99.24	93.14
Testing set (30%)					
C-1	98.33	93.94	91.18	99.25	92.54
C-2	98.00	88.57	93.94	98.50	91.18
C-3	99.33	95.65	95.65	99.64	95.65
C-4	98.33	100.00	84.38	100.00	91.53
C-5	97.67	90.62	87.88	98.88	89.23
C-6	99.33	92.86	100.00	99.27	96.30
C-7	99.33	96.67	96.67	99.63	96.67
C-8	98.00	90.32	90.32	98.88	90.32
C-9	98.67	91.43	96.97	98.88	94.12
C-10	99.67	96.15	100.00	99.64	98.04
Average	98.67	93.62	93.70	99.26	93.56

Table 6: Comparative *accu_y* examination of FHODL-AR model on UCF Sport dataset

Methods	Accuracy (%)
FHODL-AR	99.10
RNN model	97.60
DBN model	97.74
CNN model	98.26
Bi-LSTM model	97.01
GRU model	98.21

**Figure 8:** *Accu_y* analysis of FHODL-AR model with existing models on UCF Sport dataset

5 Conclusion

In this article, a new FHODL-AR technique has been developed for activity recognition on HCI driven usability. In the presented FHODL-AR technique, the input images are investigated for the identification of different human activities. For feature extraction, a modified SqueezeNet model is introduced by the inclusion of few bypass connections to the SqueezeNet among Fire modules. Besides, the FHO algorithm is utilized as a hyperparameter optimization algorithm, which in turn boosts the classification performance. To detect and categorize different kinds of activities, PNN classifier is applied. The experimental validation of the FHODL-AR technique is tested using benchmark datasets, and the outcomes reported the improvements of the FHODL-AR technique over other recent approaches with maximum accuracy of 99.10%. In the future, the detection performance of the FHODL-AR technique can be boosted by the use of ensemble fusion approaches.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] A. Jorio, S. El Fkihi, B. Elbhiri and D. Aboutajdine, "An energy-efficient clustering routing algorithm based on geographic position and residual energy for wireless sensor network," *Journal of Computer Networks and Communications*, vol. 2015, no. 8, pp. 1–11, 2015.

- [2] F. Ren and Y. Bao, "A review on human-computer interaction and intelligent robots," *International Journal of Information Technology & Decision Making*, vol. 19, no. 1, pp. 5–47, 2020.
- [3] J. Preece, Y. Rogers and H. Sharp, *Interaction Design: Beyond Human-Computer Interaction*, 4st. ed., Chichester: Wiley, 2015.
- [4] P. Pareek and A. Thakkar, "A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 2259–2322, 2020.
- [5] P. Antonik, N. Marsal, D. Brunner and D. Rontani, "Human action recognition with a large-scale brain-inspired photonic computer," *Nature Machine Intelligence*, vol. 1, no. 11, pp. 530–537, 2019.
- [6] S. Majumder and N. Kehtarnavaz, "A review of real-time human action recognition involving vision sensing," *Real-Time Image Processing and Deep Learning*, United States, Vol. 11736, pp. 9, 2021.
- [7] F. Gu, M. H. Chung, M. Chignell, S. Valae, B. Zhou *et al.*, "A survey on deep learning for human activity recognition," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–34, 2022.
- [8] K. Verma and B. Singh, "Deep multi-model fusion for human activity recognition using evolutionary algorithms," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 2, pp. 44, 2021.
- [9] P. Shrivastava, K. Singh and A. Pancham, "Classification of grain s and quality analysis u sing deep learning," *International Journal of Engineering and Advanced Technology*, vol. 11, no. 1, pp. 244–250, 2021.
- [10] Y. Abdulazeem, H. M. Balaha, W. M. Bahgat and M. Badawy, "Human action recognition based on transfer learning approach," *IEEE Access*, vol. 9, pp. 82058–82082, 2021.
- [11] M. Ronald, A. Poulouse and D. S. Han, "iSPLInception: An inception-resnet deep learning architecture for human activity recognition," *IEEE Access*, vol. 9, pp. 68985–69001, 2021.
- [12] 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," *IEEE Access*, vol. 10, pp. 47051–47062, 2022.
- [13] T. Lv, X. Wang, L. Jin, Y. Xiao and M. Song, "A hybrid network based on dense connection and weighted feature aggregation for human activity recognition," *IEEE Access*, vol. 8, pp. 68320–68368, 2020.
- [14] M. G. A. Komang, M. N. Surya and A. N. Ratna, "Human activity recognition using skeleton data and support vector machine," *Journal of Physics: Conference Series*, vol. 1192, pp. 1–9, 2019.
- [15] A. A. Yilmaz, M. S. Guzel, E. Bostanci and I. Askerzade, "A novel action recognition framework based on deep-learning and genetic algorithms," *IEEE Access*, vol. 8, pp. 100631–100644, 2020.
- [16] S. K. Challa, A. Kumar and V. B. Semwal, "A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data," *The Visual Computer*, vol. 33, no. 12, pp. 1529, 2021.
- [17] Y. Yang, R. Yang, L. Pan, J. Ma, Y. Zh *et al.*, "A lightweight deep learning algorithm for inspection of laser welding defects on safety vent of power battery," *Computers in Industry*, vol. 123, no. 4, pp. 103306, 2020.
- [18] M. Azizi, S. Talatahari and A. H. Gandomi, "Fire Hawk Optimizer: a novel metaheuristic algorithm," *Artificial Intelligence Review*, vol. 376, no. 1, pp. 113609, 2022.