Quantum behaved intelligent variant of gravitational search algorithm with deep neural networks for human activity recognition

Sonika Jindal [1],* Monika Sachdeva [2], Alok K. S. Kushwaha [3]
[1,2,3] *Dept. of Computer Science and Engineering*
[1,2] *I. K. Gujral Punjab Technical University, Jalandhar, India*
[3] *Guru Ghasidas Vishwavidyalaya , Bilaspur, India*
**Corresponding author: sonikajindal@sbsstc.ac.in*

Abstract

Human activity recognition (HAR) encompasses the detection of daily routine activities to advance usability in detecting crime and preventing dangerous activities. The recognition of activities from videos and image sequences with higher exactitude is a major challenge due to system complexities. The efficient feature optimization approach can reduce system complexities by removing ineffective features, which also improves the activity recognition performance. This research work presents a novel quantum behaved intelligent gravitational search algorithm to optimize the features for human activity recognition. The proposed intelligent variant is termed as INQGSA, which optimizes the features by using the advantageous attributes of quantum computing (QC) and intelligent gravitational search algorithm (INGSA). In INQGSA, the intelligent factor avoids the trapping of mass agents in later iterations by using the information of the best and worst agents to update the position of agents. The addition of quantum computing based attributes (such as quantum bits, their superposition, and quantum gates, etc.) ensures a better diversity of discrete optimized features. To analyze the superiority of INQGSA, the feature optimization is also conducted with the gravitational search algorithm (GSA) and the quantum-inspired binary gravitational search algorithm (QBGSA). Finally, the optimized selected features are utilized by the deep neural networks (DNN) of ResNet-50V2 and ResNet-101V2 for the classification of activities. The activity recognition experiments are conducted on the UCF101 and HMDB51 datasets. The performance comparison of the proposed HAR system with state-of-the-art techniques signifies that the proposed system is superior and effective in detecting the different activities.

Keywords: Deep neural networks; feature optimization; gravitational search algorithm; human activity recognition; quantum computing

1. Introduction

The concept of video-based HAR has aroused the interest of industrialists and academicians in developing intelligent recognition systems. The effective recognition of activities can fulfil the future needs of building smart homes and intelligent monitoring systems. The data captured as RGB videos with the cameras is an effective means of recognizing the activities with great ease. The digitization of the world has increased the use of cameras in daily life with their already existing presence in public places such as airports, banks, hospitals, etc. Moreover, human beings themselves generate a massive amount of video content and upload it on social networking and other online websites. The primary motivation for adapting to video-based human activity recognition is the availability of a wide range of applications (Xu *et al.*, 2013; Serpush & Rezaei, 2020; Özyer *et al.*, 2021).

At the early stage of human activity recognition, the researchers focused on recognizing the simple kinematic activities from a video with a plain background. Recently, the focus of researchers has turned towards the determination of activities in real-time and uncontrolled environments. Chen *et al.* (2012) used the extreme learning model for human activity recognition, which was a device displacement free recognition model. Shieh & Huang (2012) adapted a pattern recognition model to take care of aged people with video surveillance. Moreover, an autonomous falling detection algorithm was utilized to determine the falling activities. Khemchandani & Sharma (2016) proposed the robust least square twin support vector machine (RLSTW-SVM) model along with the feature descriptors of optic flow and silhouette. The work was effective in handling the heteroscedastic noise and incorporating the outlier effect. Kushwaha *et al.* (2017) used contour-based pose features from silhouettes as well as features based on the rotation invariant local binary approach.The activity classification was conducted using the multi-class support vector machine. Ijjina & Chalavadi (2017) utilized the deep convolutional neural networks with features through depth stream videos and RGB motion streams. The presented model could tolerate robust noise. Bouachir *et al.* (2018) used different machine learning and ensemble methods to determine the suicide attempt activities. The authors determined the SVM-RBF (SVM with radial basis kernel) method as superior among others. Kong *et al.* (2019) explored the three-stream convolutional neural network to determine the multi-view falling activities. The first two streams of the model adapted the Silhouettes and motion history images as the input, and the third stream considered the dynamic images. The method lacked effectiveness due to inefficient results for the lousy representation of video clips. Jaouedi *et al.* (2020) used the Gated Recurrent Neural Network for the recognition of human activities. The Kalman Filter and Gaussian Mixture Model were used to extract the features to recognize normal and sports activities. Verma *et al.* (2020) used RGB and skeleton information as the feature attributes to recognize human activities. The combined approach of convolutional and recurrent neural networks was adapted by the authors.

Although different methods are used in the discussed contributions for activity recognition, the usability of machine learning (Khan *et al.*, 2016) and deep learning (Al-Hmouz, 2020) techniques can be majorly noticed. Simple activities with fixed backgrounds can be easily recognized with higher recognition accuracy. The recognition of activities with diverse backgrounds, performed by different individuals, is a complex task. In addition, it is considerably more challenging to build automated systems with better precision. Computer vision has been used to make many automated systems, but the current systems cannot recognise very complicated human actions.

Most of the existing systems have adapted the autonomous approach to extract and select the features for activity recognition, which is less effective. The different types of activities captured in unconstrained scenarios need their relevant feature attributes to determine the type of activity. The present work has adapted distinct strategies for the different modules of the human activity recognition process. The proposed HAR system is described in four major modules: pre-processing, feature extraction, feature

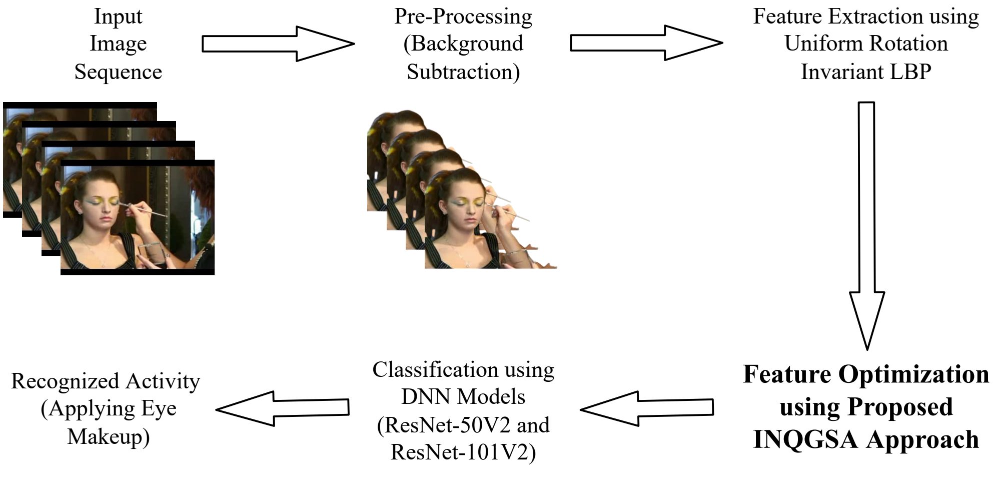


Fig. 1. Architecture of proposed HAR system.

selection, and classification. The architecture of the proposed HAR system is illustrated in Figure 1.

In the proposed HAR system, the process of recognising the activities begins with the pre-processing module that segments the background region from the extracted video frames. The processed frames (images) are evaluated to determine the features using the uniform rotation invariant LBP (Local Binary Pattern) technique. The extracted features need to be optimized to reduce the feature dimensionality and computation time by eliminating redundant and irrelevant features. Here, the INQGSA is proposed for the feature set optimization. The DNN models of ResNet-50V2 and ResNet-101V2 use the selected features to classify the activities. The performance of the proposed system is accessed for the UCF101 and HMDB51 datasets. The UCF101 dataset consists of 101 different activities, and the HMDB51 dataset is composed of 51 different activities. The focused section of the paper is the proposal of INQGSA for the feature optimization which selects the discrete feature set by adapting the attributes of an intelligent variant of GSA and the quantum computing concepts. In summary, the key contributions of the work are described as follows:

- The proposal of a novel INQGSA approach to optimize the features for the application of human activity recognition. The INQGSA approach avoids the trapping of mass agents in local optima by intelligently incorporating the advantageous attributes of QC and INGSA.
- The incorporation of advanced techniques of uniform rotation invariant LBP for multi-pose feature extraction and Deep Neural Networks (ResNet-50V2 and ResNet-101V2) for human activity recognition.
- The extensive experiments of the proposed HAR system for the video-based datasets of UCF101 and HMDB51.

The organization for the rest of the paper is described as follows. Section 2 presents the work related to feature selection and optimization techniques for human activity recognition. Section 3 illustrates the video data processing and feature extraction modules for activity recognition. Section 4 discusses the proposed INQGSA approach for the optimization of features. Section 5 exhibits the classification modules of the activities using DNN models. Section 6 describes the results and discussion of the experiments on the UCF101 and HMDB51 datasets. Section 7 concludes the paper with some future viewpoints.

2. Related work

The automation of the HAR from videos is an imperative research domain in pattern recognition as it is essential to meet the demand for a smart future in terms of automated video surveillance and smart homes. But the selection/optimization of features is the major concern in pattern recognition. During the feature extraction phase, the feature extractor can extract the different types of features for activity recognition. But the increasing feature vector can grow the dimensions of the Eigen vector, which increases the computational complexities and time consumption. Therefore, the selection and optimization of features is essential as it can determine higher recognition accuracy by consuming the least but appropriate features. The feature optimization phase contributes the relevant selected features to the HAR by removing the redundant and irrelevant features. The section describes the feature optimization based on relevant studies for video-based human activity recognition.

The feature optimization improves the HAR system performance compared to the usability of the entire feature set (Wang *et al.*, 2016). Siddiqi *et al.* (2014) presented the method of stepwise linear discriminant analysis for feature selection, which evaluates the localized features from video frames. The method was determined as efficient for the experiments on the single subject based dataset, but it lacks for the experiments on the real-time datasets having different subjects for different activities. Fang *et al.* (2014) used the inter-class distance method for feature selection and neural networks with a back propagation algorithm for activity recognition. The authors tested the results by incorporating six different feature sets and a recognition method that was evaluated as efficient compared to the Hidden Markov Model and Naive Bayes algorithm. Zheng (2015) adapted a hierarchical feature selection approach along with the classifiers of Naive Bayes and Least Squares Support Vector Machine for human activity recognition. The authors defined the requirement to place the sensors at the correct place to determine the activity accurately. Mazaar *et al.* (2016) explored the ensemble learning model by incorporating the methods of random forest and gradient boosting for feature optimization. The classification of the activities is performed using support vector machine with a linear kernel. Baldominos *et al.* (2017) conducted the feature optimization at the dimension level and attribute level using the genetic algorithm. The authors presented four different feature selection methods by incorporating with and without feature sensibility for both the dimension and attribute levels. Wang *et al.* (2018) optimized the features using the correlation-based binary particle swarm optimization approach. In this approach, the k-nearest neighbor method was used as a fitness method to determine the performance of the optimized feature set. Siddiqui *et al.* (2018) presented a codebook-based feature selection approach that includes models of visual vocabulary learning, quantization of features based on learned visual vocabulary, and representation of images by using the frequency of visual words. In the final module, activity classification was conducted using the support vector machine algorithm.

Furthermore, Siddiqi *et al.* (2019) used a normalized mutual information-based feature selection technique for the optimization of features. The authors also used linear discriminant analysis to reduce the feature space for the extracted features by using the curvelet transform. The final classification of features was performed using a hidden Markov model. Sharif *et al.* (2019) explored strong correlation and the Euclidean distance method to select the optimal feature for activity recognition. Berlin & John (2020) used a particle swarm optimization approach with a multi-objective function to reduce the feature space by selecting an appropriate feature set. The activity recognition was conducted using a deep learning neural network model. Helmi *et al.* (2021) proposed a hybrid approach of Grey Wolf Optimizer (GWO) and Gradient-Based Optimizer (GBO) for feature optimization. The GWO method was used to optimize the performance of the GBO algorithm. Tian *et al.* (2021) presented a feature selection methodology by combining the wrapper and filter feature selection approach. In this method, the initial feature selection was conducted using a game-theory filter approach, and further reselection was performed using the wrapper approach of the binary firefly algorithm. Fan & Gao (2021) integrated the deep Q-network with bee swarm optimization for the feature optimization. The bee swarm optimization approach retains the exploration and exploitation balance in the feature space, and the deep Q-network uses the advantageous attributes of reinforcement learning to make the local search space more efficient. Bulbul *et al.* (2022) focused on enhancing the performance of 3D auto-correlation gradient features. The space-time auto-correlation of gradients descriptor was used to obtain the three vectors in the method. Siddiqi & Alsirhani (2022) employed the mutual information algorithm for feature selection. The method was the extension of the max-relevance and min-redundancy to select the more appropriate and relevant features for activity recognition. In the future, the authors indicated testing the presented method in a real-time scenario.

As per the existing studies, the feature optimization techniques significantly contribute to improving the system accuracy in HAR. However, the higher recognition accuracy requires the use of an appropriate technique that can select relevant features without redundancy and can reduce the computational complexities. In addition, the standard and individual optimization techniques are observed with lacking feature attributes that increase computational cost due to higher feature dimensionality (Helmi *et al.*, 2021). The improved and ensemble approaches are essential to increase the feature optimization ability in HAR. The current work proposes the INQGSA approach, which ensembles the attributes of the quantum computing concept with an intelligent gravitational search algorithm for feature optimization. To determine the superiority of the proposed INQGSA approach, the feature optimization is also performed using the standard GSA (Rashedi *et al.*, 2009) and QBGSA (Ibrahim *et al.*, 2012). The GSA and QBGSA use $Kbest$ agents to maintain the balance of exploration and exploitation, but the $Kbest$ is a reducing function, so its value decreases over time and iterations. This decreasing value leads to the trapping of agents at later iterations. The proposed INQGSA approach overcomes this drawback by using an intelligent variant of GSA in which the position of agents is updated intelligently by using the worst ($gWorst$) and best ($gBest$) information values of the agents (Mittal & Saraswat, 2019). The mass agents get attracted towards the $gBest$ information to attain the best position and start getting away from the $gWorst$. This avoids the trapping of agents in local optima and optimizes the features effectively.

3. Video data processing and feature extraction

The section describes the pre-processing and feature extraction modules of the HAR process. These are the initial and essential modules for activity recognition.

3.1 Pre-processing

The pre-processing module segments the background region from the foreground of the image sequence (video frames). The image sequences for the proposed system are processed with a statistical model, which evaluates the variance to analyze the absolute variations and co-variance to determine the relative variations of the pixels (Singh *et al.*, 2019).

For an array of frames ($\eta_i = (\phi, \psi)$) with a starting value of SF and an ending value of EF, the variance (Var) is determined with Equation (1), and the co-variance ($Cov(\alpha, \beta)$) between the frames α and β is evaluated with Equation (2).

$$Var = \left(\frac{1}{EF} \sum_{i=0}^{EF-1} (\eta_i - \bar{\eta})^2 \right) \tag{1}$$

$$Cov(\alpha, \beta) = \left(\frac{1}{EF} \sum_{i=0}^{EF-1} \alpha_i \beta_i \right) - \left(\frac{1}{EF} \sum_{i=0}^{EF-1} \alpha_i \right) \left(\frac{1}{EF} \sum_{j=0}^{EF-1} \beta_j \right) \tag{2}$$

Where, $0 \le i < EF$ and $\bar{\eta}$ is the mean of all the frames.

The variation in the intensity of the pixel compared to other pixels is evaluated based on the co-variance between frames. The variance and co-variance values for all the pixels are stored in the reference image $Ref(\phi, \psi)$. The objects are differentiated based on the reference image.

Further, the background model is updated to incorporate the change in intensity value and background of the different frames. Exceeding the threshold value of the counter ρ indicates the requirement to update the background model. The change in the background model is performed with Equation (3).

$$Ref_{new}(\phi, \psi) = (1 - \mu) \times frame_\rho(\phi, \psi) + \mu \times Ref(\phi, \psi) \tag{3}$$

Where, $Ref_{new}(\phi, \psi)$ denotes the updated model. The symbol μ describes the updating speed, and $frame_\rho(\phi, \psi)$ depicts the current frame of the video.

3.2 Feature extraction

The features are extracted using the uniform rotation invariant LBP (Local Binary Pattern) technique from the pre-processed image sequence. The incorporation of the uniform rotation invariant is conducted to handle the activities that possess multi-view poses. The image sequences are initially converted into grayscale images to extract the features. In an LBP operator, the features of an image $I(x, y)$ with g_c as the gray level of the central pixel and g_p as the gray level of its neighbor pixels can be extracted using Equation (4) (Pietikäinen *et al.*, 2011).

$$LBP_{P,D}(x_c, y_c) = \sum_{p=0}^{p-1} s(g_p - g_c) 2^p \tag{4}$$

Where, P is the set of sample pixels in the circular neighborhood of the central pixel with radius D, $p = 0, 1, \ldots, (P-1)$, and 2^p is adapted to determine the size of histograms for the LBP operator. The values of $s(z)$ can be determined as described in Equation (5).

$$s(z) = \begin{cases} 1, & z \ge 0 \\ 0, & z < 0 \end{cases} \tag{5}$$

The local circular neighbor pixels around the central pixel with a radius of D are described in Figure 2. Here, only the uniform patterns (U) of the LBP code are incorporated to retain the statistical robustness. The uniform patterns hold the transition from 0 to 1, and the mapping of uniform LBP patterns produces $P(P-1)+3$ labels for the P sampling points.

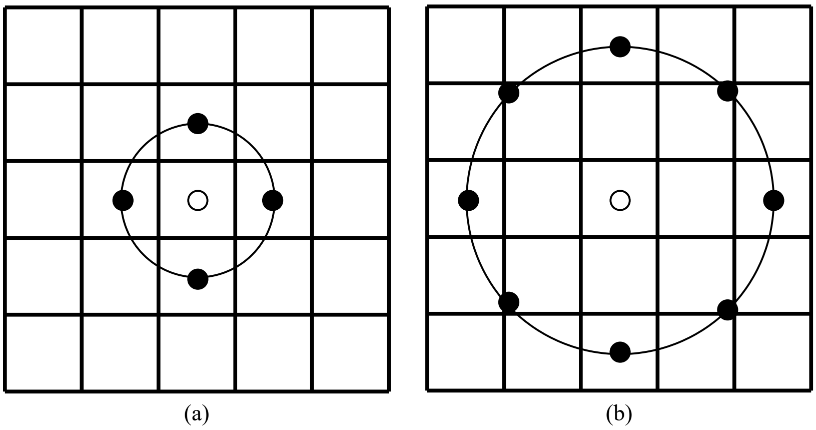


Fig. 2. Circular neighbors for central pixels in format (P,D): (a). (4,1), (b). (8,2).

With the rotation of the image $I(x,y)$, the LBP patterns are translated to another location for the rotation around their origin. The rotation of the patterns can be normalized with the rotation invariant mapping in which the LBP binary code is rotated to the minimum possible value, as depicted in Equation (6).

$$LBP_{P,D}^{ri} = \min_{i} ROR(LBP_{P,D}, i) \tag{6}$$

Where, $ROR(LBP_{P,D}, i)$ is the circular bit-wise rotation with i steps.

The features with the uniform rotation invariant LBP operator are extracted using Equations (7)- (8) that retain the robustness and higher stability (Singh *et al.*, 2019).

$$LBP_{P,D}^{riu2} = \begin{cases} \sum_{p=0}^{p-1} s\left(g_p - g_c\right), \text{ if } U(LBP_{P,D} \leq 2) \\ \quad P+1, \text{ otherwise} \end{cases} \tag{7}$$

where,

$$U(LBP_{P,D}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{p-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{8}$$

The value of $s(z)$ is evaluated using Equation (5). The uniform operator $U(LBP_{P,D})$ is a rotation invariant operator with varying bits of 0 and 1 in circular symmetry.

4. Feature optimization using proposed INQGSA approach

The optimization of features is essential for the classifier to improve the performance of the system. The present work proposes the INQGSA approach to optimize the features for human activity recognition. The GSA is a population-based meta-heuristic algorithm inspired by the physics-based Newton′s laws of motion and gravity to optimize the solution set for high dimensional problems (Rashedi *et al.*, 2009). The proposal of the INQGSA approach is presented as the standard GSA (Rashedi *et al.*, 2009) and QBGSA (Ibrahim *et al.*, 2012) algorithms lack feature optimization. In the meta-heuristic algorithm, the balance of exploration and exploitation is essential for optimization. The GSA and QBGSA use the $Kbest$ agents to retain this balance, but the value of $Kbest$ decreases with the increasing iterations because $Kbest$ is a reducing function. This decreasing value leads to the trapping of agents at later iterations. The proposed INQGSA approach adapts the advantageous attributes of QC and intelligent variant of GSA to tackle the trapping of agents. In the proposed INQGSA approach, the position of agents is updated intelligently by using the worst ($gWorst$) best ($gBest$) information values of the agents (Mittal & Saraswat, 2019). The mass agents get attracted towards the $gBest$ information to attain the best position and start getting away from the $gWorst$. This avoids the trapping of agents in local optima and optimizes the features effectively.

The proposed INQGSA algorithm begins by considering an n-dimensional system having N mass agents in which the position of the i^{th} agent can be defined by Equation (9).

$$X_i = (x_i^1, x_i^2, \dots, x_i^d, \dots, x_i^n);\; i = 1, 2, 3, \dots, N \tag{9}$$

Where, x_i^d is the position of the i^{th} mass agent in d^{th} dimension.

The force acting by the considered i^{th} agent on the j^{th} agent is determined by Equation (10).

$$F_{ij}^d(t) = G(t)\frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij}(t) + \varepsilon}(x_j^d(t) - x_i^d(t)) \tag{10}$$

Where, ε is a constant and the masses are considered as active mass (M_{aj}) and passive mass (M_{pi}) for the j^{th} agent and i^{th} agent, respectively. In Equation (10), the distance R is incorporated instead of R^2 (in law of gravity) due to better performance with only R as per the existing studies (Rashedi *et al.*, 2009). Here, the Euclidean distance R_{ij} is determined by Equation (11).

$$R_{ij}(t) = \|X_i(t), X_j(t)\|_2 \tag{11}$$

Further, the addition of stochastic attributes changes the force evaluation with the total force acting on the agent i as depicted in Equation (12). By considering the total force, the acceleration evaluation is depicted in Equation (13).

$$F_i^d(t) = \sum_{j=1, j\neq i}^{N} rand_j F_{ij}^d(t) \tag{12}$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \tag{13}$$

Where, $rand_j$ is the random number that lies in [0, 1] and M_{ii} indicates the inertial mass.

Further, the movement of the particles is determined by evaluating the change in position, velocity, and masses by Equation (14)- (16).

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{14}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{15}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \tag{16}$$

Where, $M_i = M_{ii} = M_{pi} = M_{ai}$ as the inertial and gravitational masses are assumed to be equal and calculated by the fitness function $fit_i(t)$. In Equation (16), $m_i(t)$ is evaluated using Equation (17).

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{17}$$

In the current research work, feature optimization is a minimization problem as it needs to minimize feature dimensionality. For the minimization problem, the values of $best(t)$ and $worst(t)$ are evaluated by Equations (18) and (19).

$$best(t) = min_{j\epsilon\{1,2,\dots,N\}} fit_j(t) \tag{18}$$

$$worst(t) = max_{j\epsilon\{1,2,\dots,N\}} fit_j(t) \tag{19}$$

In GSA, the mass agents can be trapped in later iterations, which can be avoided by introducing the intelligent variant of GSA. The INGSA incorporates the worst ($gWorst$) and best ($gBest$) information values of the agents to update the position of each agent intelligently (Mittal & Saraswat, 2019). For the current feature optimization problem, which is a minimization problem, the values of $gBest$ and $gWorst$ are evaluated by Equations (20)- (21).

$$gBest(t) = x_e(t) \tag{20}$$

$$gWorst(t) = x_s(t) \tag{21}$$

Where, the notations e and s are concerned with the minimum and maximum fitness functions of the intelligent mass agents, which are evaluated by Equations (22)- (23).

$$fit_e(t) = \min \{fit_1, fit_2, fit_3, \ldots, fit_N\} \tag{22}$$

$$fit_s(t) = \max \{fit_1, fit_2, fit_3, \ldots, fit_N\} \tag{23}$$

The mass agents get attracted towards the $gBest$ information to attain the best position and start getting away from the $gWorst$. The update in the position of the mass agents as per INGSA is determined by Equation (24).

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) + b(t) \times \frac{\left(gBest^d(t) - x_i^d(t)\right)}{\left|\omega \times gWorst^d(t) - x_i^d(t)\right|} \tag{24}$$

In Equation (24), the intelligent component is the third term. Here, $b(t)$ is a number that lies in [0,1] and is determined randomly. ω possesses a constant value of 0.7 and is incorporated to reduce the effect of $gWorst$ as it tries to mitigate the movement of mass agents towards $gBest$ (Mittal & Saraswat, 2019). As the mass agents move towards the $gBest$, the agents$'$ distance increases from $gWorst$ which helps to reduce the step size and avoid the trapping of agents in the local optima. Another scenario of greater distance from $gBest$ allows the agents to explore more.

Further, the concept of quantum computing is introduced with the INGSA. In quantum computing, the position and velocity of mass agent changes to quantum states with a probabilistic illustration (Ibrahim *et al.*, 2012). The Q-bit (quantum bit) is considered as the smallest unit and its state can be either 0 or 1 or their superposition, which can be analyzed for any complex numbers (C_1 and C_2) by Equation (25).

$$|\psi\rangle = C_1 |0\rangle + C_2 |1\rangle \tag{25}$$

The complex numbers C_1 and C_2 are the probability amplitudes for binary numbers 0 and 1, respectively, and they assures the normalization of states to unity by following Equation (26).

$$|C_1|^2 + |C_2|^2 = 1 \tag{26}$$

The states of the Q-bits are updated by using the quantum gates. Among the eminent quantum gates of the rotation gate, NOT gate, Hadamard gate, etc., this work incorporates the rotation gate due to its effective performance in the existing studies (Ibrahim *et al.*, 2012). The solution for the INQGSA-agents through the rotation gate is presented by Equation (27).

$$U(\Delta\theta) = \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) \\ \sin(\Delta\theta) & \cos(\Delta\theta) \end{bmatrix} \tag{27}$$

Where, $\Delta\theta$ is the rotation angle for $i = 1, 2, \ldots, n$ that determines the position of the agents in terms of quantum state.

In the INQGSA approach, the movement of the quantum mass agents is determined by updating Equation (24) with the quantum movements, which is illustrated by Equation (28).

$$\theta_{ij}^d(t+1) = \theta_{ij}^d(t) + \Delta\theta_{ij}^d(t+1) + b(t) \times \frac{\left(gBest^d(t) - \theta_{ij}^d(t)\right)}{\left|\omega \times gWorst^d(t) - \theta_{ij}^d(t)\right|} \tag{28}$$

Algorithm 1: Pseudo Code of the INQGSA approach for Feature Optimization

```
Initialize the parameters of the QC concept and GSA algorithm such as as $t_{max}$, $\vartheta_0$, $\omega$, etc.
Determine the initial fitness value for the population of intelligent mass agents.
t=1; while $t < t_{max}$ do
    for $i = 1$ to $N$ do
        Evaluate the $\vartheta$ and $a$ values for the agents.
        Evaluated the $\theta_{ij}^d(t)$ and $\Delta\theta_{ij}^d(t)$ values for the agents
        Determine the information for the agents concerning the best and worst fitness
          information.
        Evaluate the fitness value.
        Update the position ($\theta_{ij}^d(t+1)$) and velocity ($\Delta\theta_{ij}^d(t+1)$) values for the agents using
          Equations (28) and (29).
    end
end
Store the optimal features determined by the best agents at optimal positions and best fitness
  value.
```

Where,

$$\Delta\theta_{ij}^d(t+1) = rand_i \times \Delta\theta_{ij}^d(t) + a_{ij}^d(t) \tag{29}$$

In Equation (29), $a_{ij}^d(t)$ is evaluated by putting the values of Equations (10)- (12) into Equation (13), which is further derived as per the INQGSA approach. In Equation (10), the value of ε is neglected as it is constant. The derived formulation for $a_{ij}^d(t)$ as per the INQGSA approach is presented by Equation (30).

$$a_{ij}^d(t) = \sum_{j=1, j\neq i} \left[rand_j \times \vartheta \times \gamma_i^k \times \left(\theta_{kj}^d(t) - \theta_{ik}^d(t) \right) \right] \tag{30}$$

Where, the symbol ϑ is $G(t)$ which decreases from ϑ_{max} to ϑ_{min} depending on the rotation angle. The ratio of the mass (M_{aj}) and distance (R_{ij}) are presented by a decision parameter (γ_i^k) which is evaluated by Equations (31)- (32) (Ibrahim *et al.*, 2012).

$$\gamma_i^k = \begin{cases} \lambda_i^k + 1, \text{if } fit(\theta_k^d(t)) = fit(\theta_i^d(t)) \\ \lambda_i^k, \text{ otherwise} \end{cases} \tag{31}$$

$$\lambda_i^k = \begin{cases} 1, \ M_k > M_i \text{ and } R_{ik} \leq \tau \\ 0, \text{ otherwise} \end{cases} \tag{32}$$

Where, τ represents the maximum number of different bits out of total bits in between the i^{th} and k^{th} agents that can put the active force on the i^{th} agent.

The optimized features are selected by the intelligent quantum mass agents upon the completion of their maximum iterations. At maximum iterations, the features selected by the best agents that possess optimized position, are retained. The pseudo-code of the feature optimization process using the INQGSA approach is illustrated in Algorithm 1.

5. Classification and recognition of activities

The classification of the activities is conducted with the deep residual networks (ResNet), which possess the deep neural network (DNN) architecture. DNN models are capable of mapping the features of layer data within deep networks. The network architecture of ResNet is a series of blocks connected to each other with parallel shortcut links for the output. The basic structure of the residual network block and its internal learning process are illustrated in Figure 3.

In Figure 3(a), incorporating the parameterized layer after the Addition module can reduce the ResNet's advantages, but incorporating the non-parameterized layer (ReLU) after the Addition has little impact on the ResNet (Kiliç *et al.*, 2020).The conventional CNN is not significant for in-depth learning as the error increases (due to over-fitting) with the increase in the depth of layers. In ResNet, the residual values are formed after adding the blocks, which are fed to the succeeding layers in the model.

In Figure 3(b), x is incorporated as an input, and the output is obtained after the ReLU operation in the form of $H(x) = F(x) + x$. Here, the input (x) is passed from the weight layer (w), and the results are acquired in the form of $F(x)$. The final output is determined by adding the x input to $F(x)$.

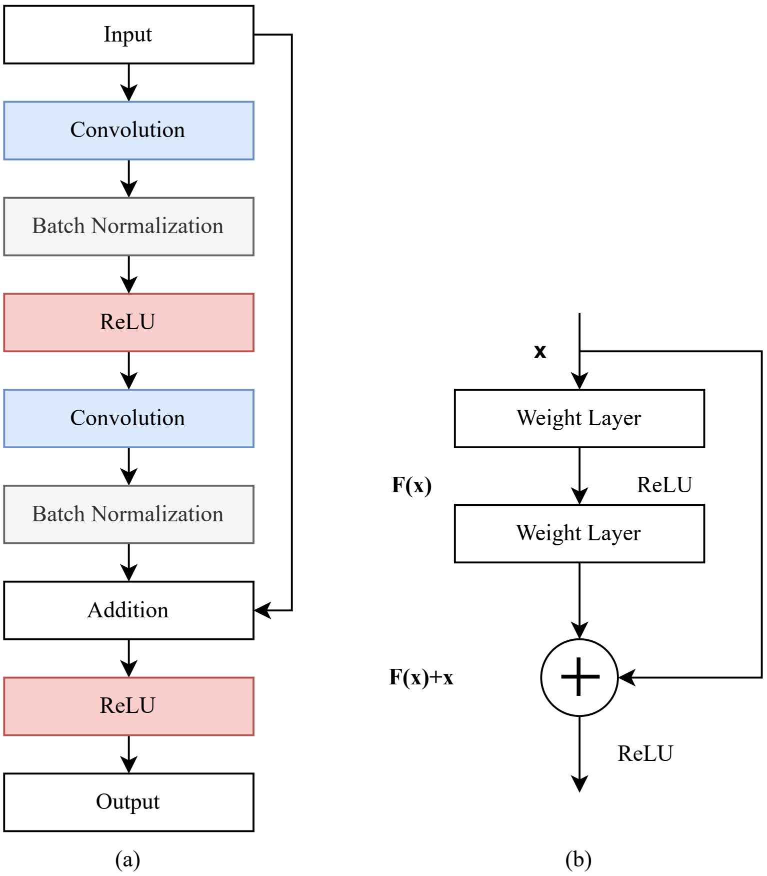


Fig. 3. (a) Basic ResNet Block (b) Internal Learning Process of Residual Block.

In this research work, the ResNet with 50 and 101 layers is adapted for the classification of activities. These networks are constructed using the architecture of 3-layer bottleneck blocks. There are 3.8×10^9 and 7.6×10^9 FLOPs in ResNet-50 and ResNet-101 respectively. The complexity of these networks is lower than VGG16/19, even after increasing the deep layers. The architectures of ResNet-50 and ResNet-101 are described in Table 1.

Here, version 2 (V2) of the ResNet is incorporated to direct the identity connections from input to output by removing the last non-linearity, which enhances the learning process and hence the classification of activities. In ResNet V2, the weight layers are pre-activated instead of post-activation. The present research work has incorporated the ResNet-50V2 and ResNet-101V2 for the human activities classification.

6. Experimental results and discussion

The results for the proposed HAR system are determined using evaluation measures of precision, recall, and f-measure for the experiments on the UCF101 and HMDB51 datasets. Furthermore, the recognition accuracy is also evaluated for the quantitative analysis of the proposed HAR system. The recognition accuracy is described in Equation (33).

$$\text{Recognition Accuracy} = \frac{\text{Correctly Classified Instances}}{\text{Total Number of Instances}} \times 100 \tag{33}$$

Table 1. Layer Architecture of the Residual Networks.

Layer	ResNet-50	ResNet-101	Output Size
Convolutional 1	7×7, 64, stride 2		112×112
Convolutional 2	3×3 max pooling, stride 2 $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	56×56
Convolutional 3	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	28×28
Convolutional 4	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	14×14
Convolutional 5	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	7×7
	Average Pooling, 1000 Fully Connected Softmax		1×1

Table 2. Statistics of Datasets.

Parameter	UCF101	HMDB51
Actions	101	51
Resolution	320 × 240	320 × 240
Video Clips	13,320	6,766
Frame Rate	25 fps	30 fps
Min. Video Clip Length	1.06 sec	1 sec
Min. Video Clips Per Action	100	101

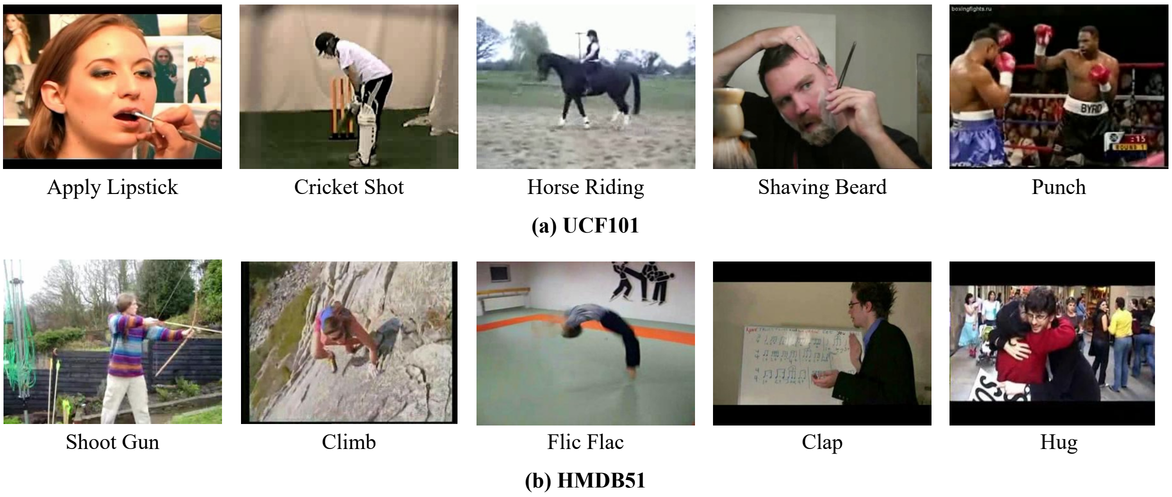
6.1 Datasets

The present work has utilized the UCF101 (Soomro *et al.*, 2012) and HMDB51 (Kuehne *et al.*, 2011) datasets, which are video-based datasets. The UCF101 dataset is composed of 101 realistic action videos gathered from YouTube. There are 13,320 videos of different actions available in this dataset, and the different activities are divided into five categories: sports, playing musical instruments, human-human interaction, body-motion, and human-object interaction. Whereas, the HMDB51 dataset is collected from the Prelinger archive, Google, and YouTube videos. The HMDB51 dataset embodies 6,766 video clips related to 51 action categories, which are majorly divided into five categories: general body movements, body movement for human interaction, body movement for object interaction, general facial actions, and facial actions with object manipulation. The statistics of both the datasets are illustrated in Table 2, and some sample frames indicating different activities are depicted in Figure 4.

6.2 Result evaluation

To perform the experiments for the proposed HAR system, both the datasets (UCF101 and HMDB51) are divided separately into the training and testing proportions of approximately 90:10. For both the datasets, 1,650 frames per activity are extracted. A total of 166,650 frames from the UCF101 dataset and 84,150 frames from the HMDB51 dataset are extracted. Among the total 1,650 frames per activity, 1,500 frames are utilized for training the residual networks and 150 frames are utilized for testing. The description of the training and testing settings is depicted in Table 3.

Before evaluating the testing results for the proposed HAR system, the data is validated by splitting the training data frames (151,500 frames of the UCF101 dataset and 76,500 frames of the HMDB51 dataset) into the ratio of 80:20. The 80% of the data (121,200 frames of UCF101 dataset and 61,200 frames of HMDB51 dataset) is utilized for the training and 20% of the data (30,300 frames of UCF101



Apply Lipstick Cricket Shot Horse Riding Shaving Beard Punch

(a) UCF101

Shoot Gun Climb Flic Flac Clap Hug

(b) HMDB51

Fig. 4. Sample Frames illustrating different Activities (a) UCF101 dataset (b) HMDB51 dataset.

Table 3. Training and Testing Setting.

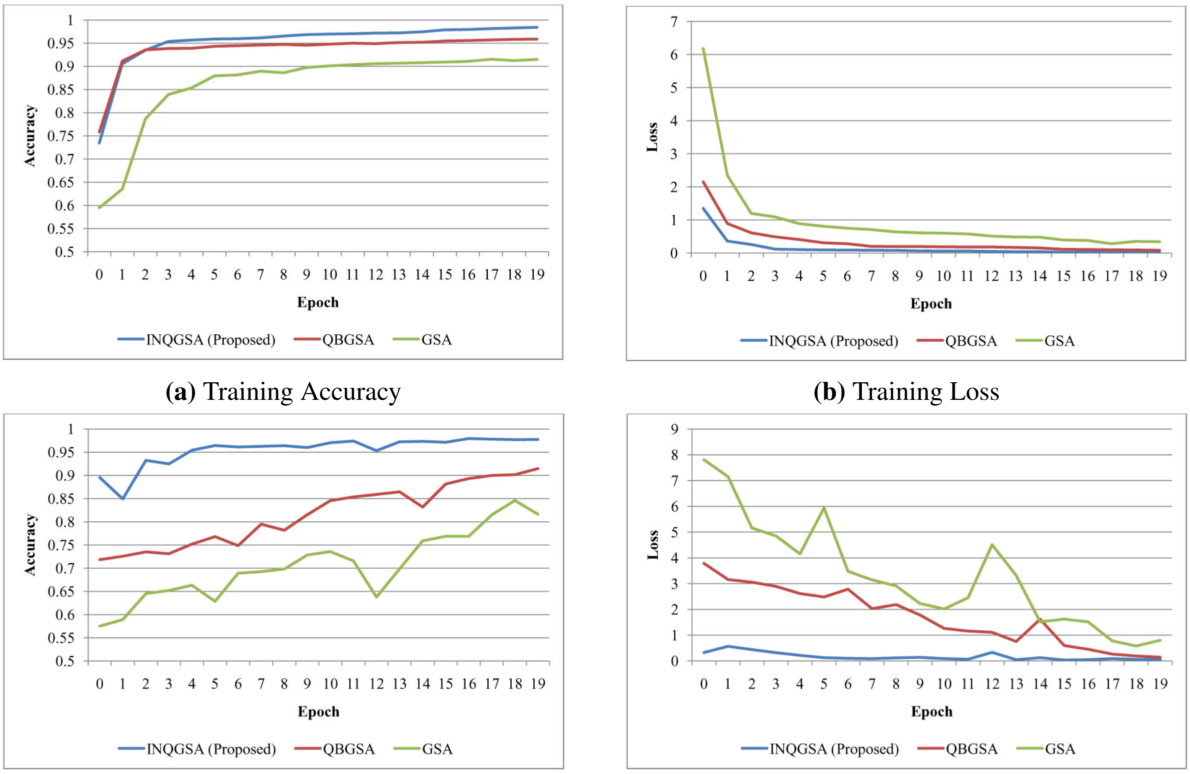
Parameter	Value
Input of spatial stream	Size of single frame = 3×224×224
Total number of frames	1,650 frames per activity
Number of frames (Training)	1,500 frames per activity
Batch Size	32
Number of Epochs	20
Initial learning rate	$5e^4$
Number of frames (Testing)	150 frames per activity

dataset and 15,300 frames of HMDB51 dataset) is used for the validation. Figures 5- 8 illustrate the accuracy and loss curves over the 20 epochs during the training and validation for both the UCF101 and HMDB51 datasets. In Figures 5- 8, the results are determined by incorporating the different feature optimization techniques (GSA, QBGSA, and proposed INQGSA) along with the DNN classifiers of ResNet-50V2 and ResNet-101V2.

The graphs depicted in Figures 5- 8 indicate the higher oscillation of validation results in the case of GSA and QBGSA, which is due to the trapping of agents with the increase of epochs. On the other hand, the proposed INQGSA can be seen with the minor oscillations of result values. The training of the techniques can be found to be smooth compared to the validation results. The validation results for the UCF101 and HMDB51 datasets are illustrated in Tables 4 and 5, respectively. These results clearly indicate the higher accuracy and lower loss values of the proposed models. For the UCF101 dataset, the maximum validation accuracy values of 97.95% and 98.98% are attained by the proposed INQGSA+ResNet50V2 technique and the proposed INQGSA+ResNet101V2 technique, respectively. Furthermore, these values are 96.92% and 98.25% in the case of the HMDB51 dataset for the aforementioned techniques. These validation results are higher than other feature optimization techniques, which indicate the superiority of the proposed approach. It also indicates that the ResNet-101V2 attained superior performance to the ResNet-50V2.

Further, the testing results of the proposed INQGSA approach and other optimization techniques with ResNet-50V2 and ResNet-101V2 classifiers are determined in terms of precision, recall, f-measure score, and recognition accuracy. The classification results for the UCF101 and HMDB51 datasets are described in Tables 6 and 7.

From the classification results depicted in Tables 6 and 7, it can be seen that the results values of INQGSA with both the DNN models (ResNet-50V2 and ResNet-101V2) are higher than the results evaluated with QBGSA and GSA. It indicates that the INQGSA can optimize the features more efficiently



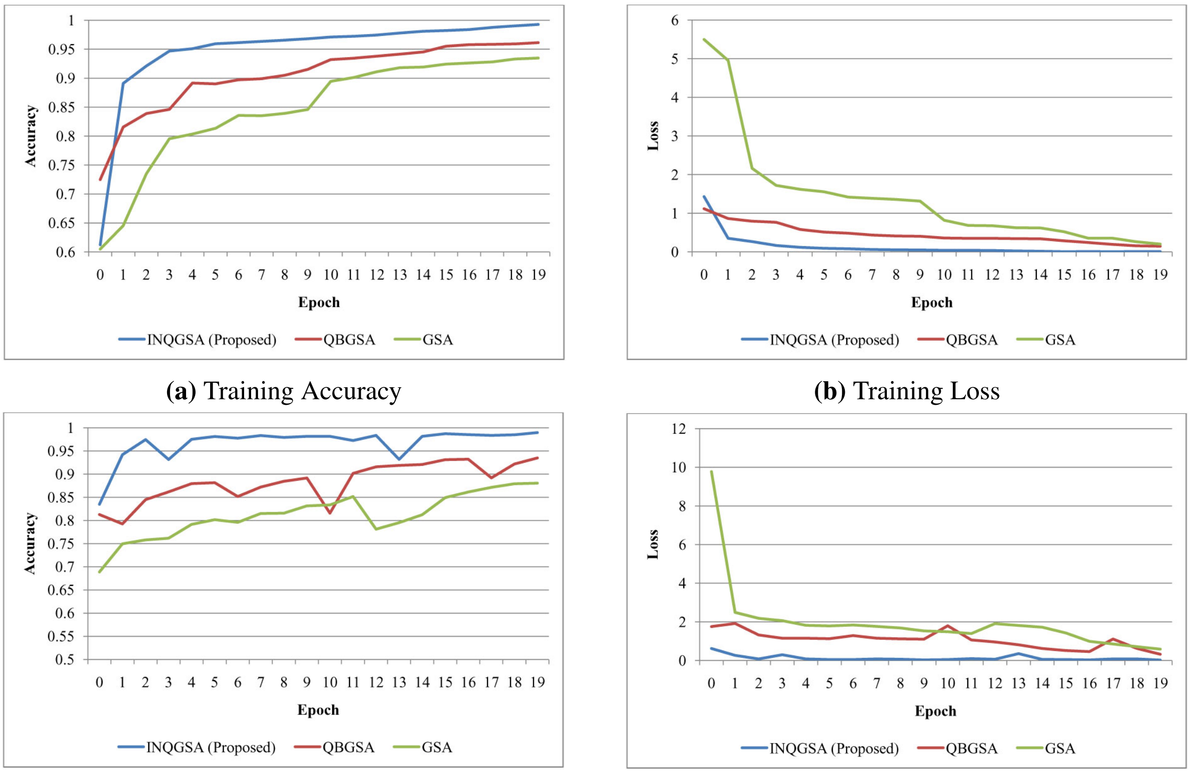
(a) Training Accuracy

(b) Training Loss

(c) Validation Accuracy

(d) Validation Loss

Fig. 5. Performance of ResNet-50V2 Classifier with different Feature Optimization Techniques for the UCF101 Dataset.



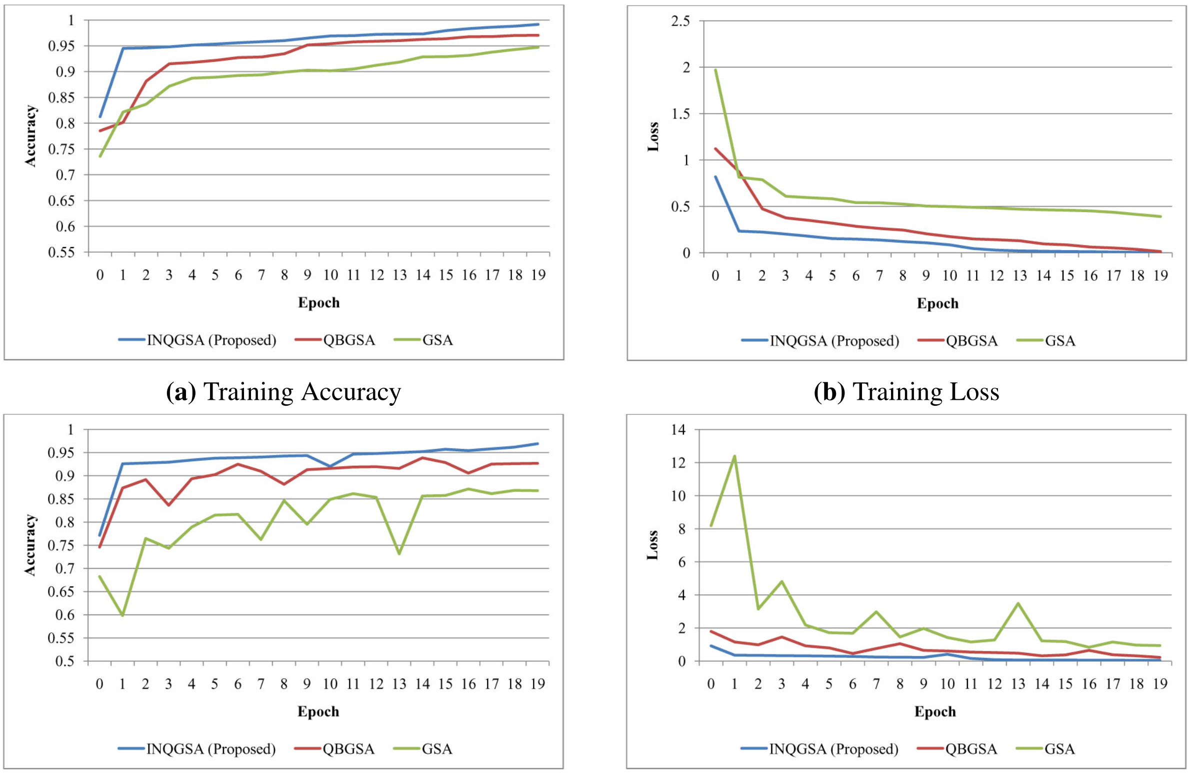
(a) Training Accuracy

(b) Training Loss

(c) Validation Accuracy

(d) Validation Loss

Fig. 6. Performance of ResNet-101V2 Classifier with different Feature Optimization Techniques for the UCF101 Dataset.



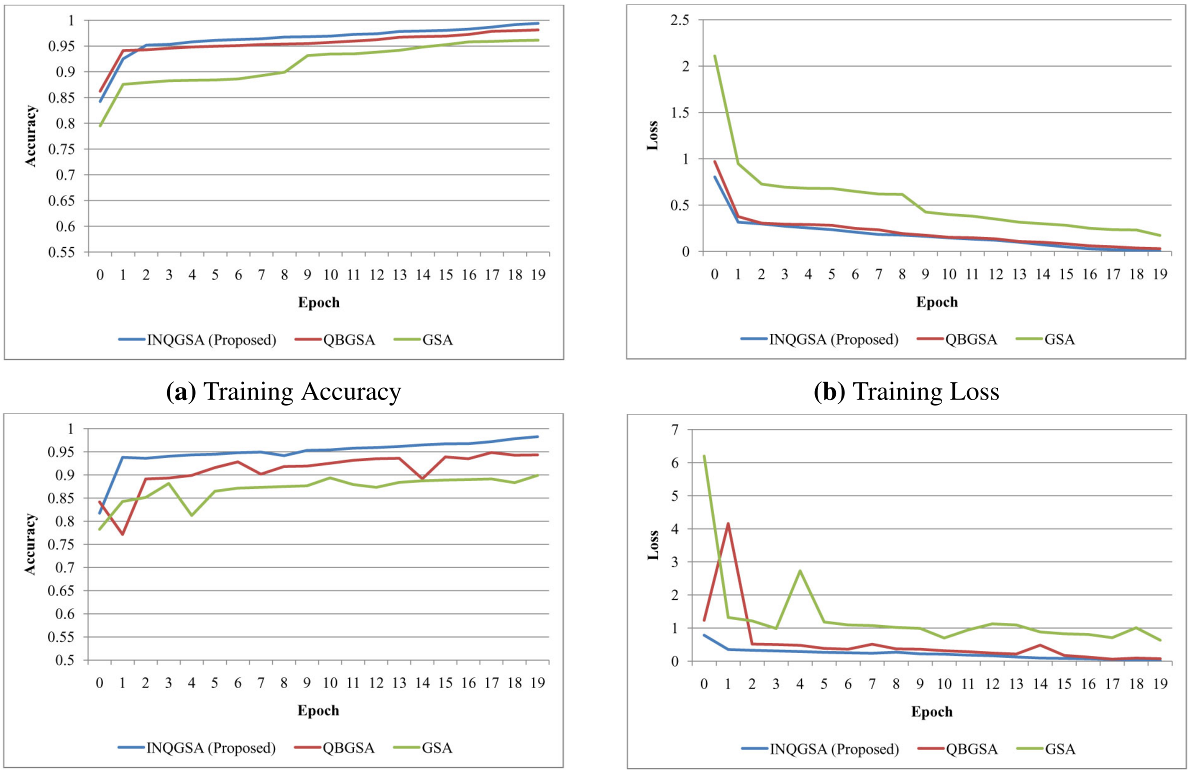
(a) Training Accuracy

(b) Training Loss

(c) Validation Accuracy

(d) Validation Loss

Fig. 7. Performance of ResNet-50V2 Classifier with different Feature Optimization Techniques for the HMDB51 Dataset.



(a) Training Accuracy

(b) Training Loss

(c) Validation Accuracy

(d) Validation Loss

Fig. 8. Performance of ResNet-101V2 Classifier with different Feature Optimization Techniques for the HMDB51 Dataset.

Table 4. Validation Results for the UCF101 Dataset.

Technique	Max. Validation Accuracy	Min. Validation Loss
GSA+ResNet50V2	84.59%	0.579
GSA+ResNet101V2	88.06%	0.581
QBGSA+ResNet50V2	91.48%	0.1431
QBGSA+ResNet101V2	93.48%	0.318
Proposed INQGSA+ResNet50V2	97.95%	0.0327
Proposed INQGSA+ResNet101V2	98.98%	0.0243

Table 5. Validation Results for the HMDB51 Dataset.

Technique	Max. Validation Accuracy	Min. Validation Loss
GSA+ResNet50V2	86.85%	0.837
GSA+ResNet101V2	89.91%	0.631
QBGSA+ResNet50V2	93.85%	0.2247
QBGSA+ResNet101V2	94.82%	0.061
Proposed INQGSA+ResNet50V2	96.92%	0.0415
Proposed INQGSA+ResNet101V2	98.25%	0.0173

compared to the GSA and QBGSA. The maximum recognition accuracy values of 96.16% and 97.11% are attained by the proposed INQGSA+ResNet101V2 technique for the UCF101 and HMDB51 datasets, respectively. As the proposed techniques are superior to other optimization techniques, therefore only the proposed techniques are incorporated for further comparison with state-of-the-art techniques.

6.3 Comparative analysis

The proposed HAR system has incorporated the RGB frames for activity recognition from video datasets. Therefore, the comparative analysis of the proposed system is conducted with most of the RGB-based techniques for the experiments on the UCF101 and HMDB51 datasets. The comparative analysis of the proposed system with state-of-the-art techniques in terms of recognition accuracy is summarized in Table 8.

The proposed INQGSA approach outperformed with both the classifiers (ResNet-50V2 and ResNet-101V2) compared to the state-of-the-art techniques. For the UCF101 and HMDB51 datasets, the proposed INQGSA+ResNet101V2 technique has attained 0.78% and 1.27% higher accuracy values than the INQGSA+ResNet50V2 technique, respectively.

For the UCF101 dataset, the recognition accuracy of the proposed INQGSA+ResNet101V2 technique is 7.06% higher than MIFS (Multi-skIp Feature Stacking) (Lan *et al.*, 2015), 4.26% than Motion Map+MIFS (Sun *et al.*, 2018), 7.26% than MiCT-Net (Mixed Convolutional Tube Network) (Zhou *et al.*, 2018), 4.66% than CNN-OFF (Xu *et al.*, 2021), 3.27% than CNN (weighted product fusion) (Singh *et al.*, 2021), 4.12% than CNN (weighted average fusion) (Singh *et al.*, 2021), 4.6% than CNN (max fusion) (Singh *et al.*, 2021), 7.49% than CNN (sum fusion) (Singh *et al.*, 2021), 7.93% than CNN (spatio-

Table 6. Classification Results for the UCF101 Dataset.

Technique	Precision	Recall	F-Measure	Recognition Accuracy
GSA+ResNet50V2	86.08%	83.14%	84.58%	83.14%
GSA+ResNet101V2	87.89%	86.23%	87.05%	86.23%
QBGSA+ResNet50V2	90.44%	89.53%	89.98%	89.53%
QBGSA+ResNet101V2	94.15%	92.77%	93.45%	92.77%
Proposed INQGSA+ResNet50V2	96.17%	95.38%	95.77%	95.38%
Proposed INQGSA+ResNet101V2	96.91%	96.16%	96.53%	96.16%

Table 7. Classification Results for the HMDB51 Dataset.

Technique	Precision	Recall	F-Measure	Recognition Accuracy
GSA+ResNet50V2	88.12%	85.76%	86.92%	85.76%
GSA+ResNet101V2	90.47%	88.35%	89.40%	88.35%
QBGSA+ResNet50V2	93.48%	92.07%	92.77%	92.07%
QBGSA+ResNet101V2	94.13%	92.97%	93.55%	92.97%
Proposed INQGSA+ResNet50V2	97.09%	95.84%	96.46%	95.84%
Proposed INQGSA+ResNet101V2	98.37%	97.11%	97.74%	97.11%

Table 8. Comparison of the Proposed HAR System with State-of-the-art Techniques.

Technique	UCF101	HMDB51
MIFS (Lan *et al.*, 2015)	89.1%	65.1%
Motion Map+MIFS (Sun *et al.*, 2018)	91.9%	73.7%
MiCT-Net (Zhou *et al.*, 2018)	88.9%	63.8%
M-SVM (Sharif *et al.*, 2019)	-	92.6%
CNN-OFF (Xu *et al.*, 2021)	91.5%	67.9%
CNN (weighted product fusion) (Singh *et al.*, 2021)	92.89%	64.13%
CNN (weighted average fusion) (Singh *et al.*, 2021)	92.04%	63.87%
CNN (max fusion) (Singh *et al.*, 2021)	91.56%	62.79%
CNN (sum fusion) (Singh *et al.*, 2021)	88.67%	62.32%
CNN (spatio-temp) (Singh *et al.*, 2021)	88.23%	61.89%
CNN (spatial) (Singh *et al.*, 2021)	82.23%	57.20%
MSM-ResNets (Zong *et al.*, 2021)	93.5%	66.7%
PDaUM+DCNN (Khan *et al.*, 2021)	-	81.4%
Proposed INQGSA+ResNet50V2	95.38%	95.84%
Proposed INQGSA+ResNet101V2	96.16%	97.11%

temp) (Singh *et al.*, 2021), 13.93% than CNN (spatial) (Singh *et al.*, 2021), 2.66% than MSM-ResNets (Motion Saliency based multi-stream Multiplier ResNets) (Zong *et al.*, 2021).

For the HMDB51 dataset, the recognition accuracy of the proposed INQGSA+ResNet101V2 technique is 32.01% higher than MIFS (Lan *et al.*, 2015), 23.41% than Motion Map+MIFS (Sun *et al.*, 2018), 33.31% than MiCT-Net (Zhou *et al.*, 2018), 4.51% than M-SVM (Multi-class Support Vector Machine) (Sharif *et al.*, 2019), 29.21% than CNN-OFF (Xu *et al.*, 2021), 32.98% than CNN (weighted product fusion) (Singh *et al.*, 2021), 33.24% than CNN (weighted average fusion) (Singh *et al.*, 2021), 34.32% than CNN (max fusion) (Singh *et al.*, 2021), 34.79% than CNN (sum fusion) (Singh *et al.*, 2021), 35.22% than CNN (spatio-temp) (Singh *et al.*, 2021), 39.91% than CNN (spatial) (Singh *et al.*, 2021), 30.41% than MSM-ResNets (Zong *et al.*, 2021), and 15.71% than PDaUM (Poisson distribution along with Univariate Measures) + DCNN (Deep Convolutional Neural Network) (Khan *et al.*, 2021).

These comparisons indicate the superiority of the results for the proposed techniques over other techniques. Although the accuracy differences between the proposed techniques and other techniques are readily visible for both the datasets, a significant improvement in the results can be observed for the HMDB51 dataset. These results demonstrate that the proposed INQGSA technique considerably enhances the features that aid in the more accurate recognition of activities.

7. Conclusion

This paper proposed the INQGSA approach to optimize the features for human activity recognition. The proposed INQGSA approach intelligently updates the position of mass agents to avoid the trapping of agents in later iterations, which occurred in GSA and QBGSA. In this work, these intelligent attributes helps to improve the feature optimization for activity recognition. In the overall human activity recognition system, a sequence of the latest techniques is incorporated for the different modules of activity recognition. The system incorporated the key techniques of uniform rotation invariant LBP for feature extraction, the proposed INQGSA approach for feature optimization, and deep neural network models (ResNet-50V2 and ResNet-101V2) for classification. The feature optimization technique reduces the complexity of the classifiers by feeding the selected features. The results of the proposed HAR system are evaluated for the UCF101 and HMDB51 datasets. For the UCF101 dataset, the proposed INQGSA+ResNet50V2 technique and the proposed INQGSA+ResNet101V2 techniques attained recognition accuracy of 95.38% and 96.16%, respectively. These values for the HMDB51 dataset are 95.84% and 97.11%, respectively. The comparative analysis of the proposed techniques with GSA and QBGSA based optimization techniques and state-of-the-art techniques indicates the outperformed performance of the proposed techniques.

In the future, the proposed INQGSA approach can be utilized for other applications such as network optimization, scheduling, robotic programs, etc. Moreover, the proposed HAR system can also be implemented in real time to determine abnormal activities in public places.

Acknowledgement

The authors gratefully acknowledge the Department of Computer Science and Engineering, I. K. Gujral Punjab Technical University, Jalandhar, India, for providing the opportunity to conduct this research.

References

Al-Hmouz, R. (2020). Deep learning autoencoder approach: Automatic recognition of artistic Arabic calligraphy types. *Kuwait Journal of Science*, **47**(3), 2-14.

Baldominos, A., Isasi, P. & Saez, Y. (2017). Feature selection for physical activity recognition using genetic algorithms. In *2017 IEEE Congress on Evolutionary Computation (CEC)* (pp. 2185-2192). Donostia, Spain: IEEE.

Berlin, S. J. & John, M. (2020). Particle swarm optimization with deep learning for human action recognition. *Multimedia Tools and Applications*, **79**(25), 17349-17371.

Bouachir, W., Gouiaa, R., Li, B. & Noumeir, R. (2018). Intelligent video surveillance for real-time detection of suicide attempts. *Pattern Recognition Letters*, **110**, 1-7.

Bulbul, M. F., Islam, S., Azme, Z., Pareek, P., Kabir, M. & Ali, H. (2022). Enhancing the performance of 3D auto-correlation gradient features in depth action classification. *International Journal of Multimedia Information Retrieval*, **11**, 61-76.

Chen, Y., Zhao, Z., Wang, S. & Chen, Z. (2012). Extreme learning machine-based device displacement free activity recognition model. *Soft Computing*, **16**(9), 1617-1625.

Fan, C. & Gao, F. (2021). Enhanced human activity recognition using wearable sensors via a hybrid feature selection method. *Sensors*, **21**(19), 6434(1-25).

Fang, H., He, L., Si, H., Liu, P. & Xie, X. (2014). Human activity recognition based on feature selection in smart home using back-propagation algorithm. *ISA Transactions*, **53**(5), 1629-1638.

Helmi, A. M., Al-Qaness, M. A., Dahou, A., Damaševičius, R., Krilavičius, T. & Elaziz, M. A. (2021). A novel hybrid gradient-based optimizer and grey wolf optimizer feature selection method for human activity recognition using smartphone sensors. *Entropy*, **23**(8), 1065(1-20).

Ibrahim, A. A., Mohamed, A. & Shareef, H. (2012). A novel quantum-inspired binary gravitational search algorithm in obtaining optimal power quality monitor placement. *Journal of Applied Sciences*, **12**(9), 822-830.

Ijjina, E. P. & Chalavadi, K. M. (2017). Human action recognition in RGB-D videos using motion sequence information and deep learning. *Pattern Recognition*, **72**, 504-516.

Jaouedi, N., Boujnah, N. & Bouhlel, M. S. (2020). A new hybrid deep learning model for human action recognition. *Journal of King Saud University-Computer and Information Sciences*, **32**(4), 447-453.

Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (pp. 1942-1948). Perth, Australia: IEEE.

Kennedy, J. & Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. In *1997 IEEE International conference on systems, man, and cybernetics. Computational cybernetics and simulation* (pp. 4104-4108). Orlando, USA: IEEE.

Khan, M. A., Zhang, Y. D., Khan, S. A., Attique, M., Rehman, A. & Seo, S. (2021). A resource conscious human action recognition framework using 26-layered deep convolutional neural network. Multimedia Tools and Applications. *Multimedia Tools and Applications*, **80**(28), 35827-35849.

Khan, W., Daud, A., Nasir, J. A. & Amjad, T. (2016). A survey on the state-of-the-art machine learning models in the context of NLP. *Kuwait journal of Science*, **43**(4), 95-113.

Khemchandani, R. & Sharma, S. (2016). Robust least squares twin support vector machine for human activity recognition. *Applied Soft Computing*, **47**, 33-46.

Kiliç, Ş., Askerzade, İ. & Kaya, Y. (2020). Using ResNet Transfer Deep Learning Methods in Person Identification According to Physical Actions. *IEEE Access*, **8**, pp.220364-220373.

Kong, Y., Huang, J., Huang, S., Wei, Z. & Wang, S. (2019). Learning spatiotemporal representations for human fall detection in surveillance video. *Journal of Visual Communication and Image Representation*, **59**, 215-230.

Kuehne, H., Jhuang, H., Garrote, E., Poggio, T. & Serre, T. (2011). HMDB: a large video database for human motion recognition. In *2011 International conference on computer vision*, (pp. 2556-2563). Barcelona, Spain: IEEE.

Kushwaha, A.K.S., Srivastava, S. & Srivastava, R. (2017). Multi-view human activity recognition based on silhouette and uniform rotation invariant local binary patterns. *Multimedia Systems*, **23**(4), 451-467.

Lan, Z., Lin, M., Li, X., Hauptmann, A. G. & Raj, B. (2015). Beyond gaussian pyramid: Multi-skip feature stacking for action recognition. In *IEEE conference on computer vision and pattern recognition*, (pp. 204-212). Boston, MA: IEEE.

Liu, L. C., Rustia, D. J. A. & Lin, T. T. (2021). Remote Surveillance Video Activity Recognition Using Spatiotemporal Convolutional Neural Networks for Greenhouse Workload Analysis. In *2021 ASABE Annual International Virtual Meeting*. American Society of Agricultural and Biological Engineers.

Mazaar, H., Emary, E. & Onsi, H. (2016). Ensemble based-feature selection on human activity recognition. In *10th International Conference on Informatics and Systems* (pp. 81-87). Giza, Egypt: ACM.

Mittal, H. & Saraswat, M. (2019). An automatic nuclei segmentation method using intelligent gravitational search algorithm based superpixel clustering. *Swarm and Evolutionary Computation*, **45**, 15-32.

Özyer, T., Ak, D. S. & Alhajj, R. (2021). Human action recognition approaches with video datasets—A survey. *Knowledge-Based Systems*, **222**, 106995 (1-36).

Pietikäinen, M., Hadid, A., Zhao, G. & Ahonen, T. (2011). Local binary patterns for still images. In *Computer vision using local binary patterns* (pp. 13-47). Springer, London.

Rashedi, E., Nezamabadi-Pour, H. & Saryazdi, S. (2009). GSA: a gravitational search algorithm. *Information Sciences*, **179**(13), 2232-2248.

Serpush, F. & Rezaei, M. (2020). Complex human action recognition in live videos using hybrid fr-dl method. arXiv preprint arXiv:2007.02811.

Sharif, A., Khan, M. A., Javed, K., Gulfam, H., Iqbal, T., Saba, T., Ali, H. & Nisar, W. (2019). Intelligent human action recognition: A framework of optimal features selection based on Euclidean distance and strong correlation. *Journal of Control Engineering and Applied Informatics*, **21**(3), 3-11.

Shieh, W. Y. & Huang, J. C. (2012). Falling-incident detection and throughput enhancement in a multi-camera video-surveillance system. *Medical Engineering & Physics*, **37**(7), 954-963.

Siddiqi, M. H., Ali, R., Rana, M., Hong, E. K., Kim, E. S. & Lee, S. (2014). Video-based human activity recognition using multilevel wavelet decomposition and stepwise linear discriminant analysis. *Sensors*, **14**(4), 6370-6392.

Siddiqi, M. H., Alruwaili, M. & Ali, A. (2019). A novel feature selection method for video-based human activity recognition systems. *IEEE Access*, **7**, 119593-119602.

Siddiqi, M. H. & Alsirhani, A. (2022). An Efficient Feature Selection Method for Video-Based Activity Recognition Systems. *Mathematical Problems in Engineering*, 5486004(1-13). DOI: 10.1155/2022/5486004.

Siddiqui, S., Khan, M. A., Bashir, K., Sharif, M., Azam, F. & Javed, M. Y. (2018). Human action recognition: a construction of codebook by discriminative features selection approach. *International Journal of Applied Pattern Recognition*, **5**(3), 206-228.

Singh, R., Khurana, R., Kushwaha, A. K. S. & Srivastava, R. (2021). A dual stream model for activity recognition: Exploiting residual-cnn with transfer learning. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, **9**(1), 28-38.

Singh, R., Kushwaha, A. K. S. & Srivastava, R. (2019). Multi-view recognition system for human activity based on multiple features for video surveillance system. *Multimedia Tools and Applications*, **78**(12), 17165-17196.

Soomro, K., Zamir, A. R. & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402.

Sun, Y., Wu, X., Yu, W. & Yu, F. (2018). Action recognition with motion map 3D network. *Neurocomputing*, **297**, 33-39.

Tian, Y., Zhang, J., Li, L. & Liu, Z. (2021). A Novel Sensor-Based Human Activity Recognition Method Based on Hybrid Feature Selection and Combinational Optimization. *IEEE Access*, **9**, 107235-107249.

Verma, P., Sah, A. & Srivastava, R. (2020). Deep learning-based multi-modal approach using RGB and skeleton sequences for human activity recognition. *Multimedia Systems*, **26**(6), 671-685.

Wang, A., Chen, G., Yang, J., Zhao, S. & Chang, C. Y. (2016). A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sensors Journal*, **16**(11), 4566-4578.

Wang, H., Ke, R., Li, J., An, Y., Wang, K. & Yu, L. (2018). A correlation-based binary particle swarm optimization method for feature selection in human activity recognition. *International Journal of Distributed Sensor Networks*, **14**(4), 1-17.

Xu, J., Song, R., Wei, H., Guo, J., Zhou, Y. & Huang, X. (2021). A fast human action recognition network based on spatio-temporal features. *Neurocomputing*, **441**, 350-358.

Xu, X., Tang, J., Zhang, X., Liu, X., Zhang, H. & Qiu, Y. (2013). Exploring techniques for vision based human activity recognition: Methods, systems, and evaluation. *Sensors*, **13**(2), 1635-1650.

Zheng, Y. (2015). Human activity recognition based on the hierarchical feature selection and classification framework. *Journal of Electrical and Computer Engineering*, 140820(1-9). DOI: 10.1155/2015/140820.

Zhou, Y., Sun, X., Zha, Z. J. & Zeng, W. (2018). Mict: Mixed 3d/2d convolutional tube for human action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 449-458). Salt Lake City, USA: IEEE.

Zong, M., Wang, R., Chen, X., Chen, Z. & Gong, Y. (2021). Motion saliency based multi-stream multiplier ResNets for action recognition. *Image and Vision Computing*, **107**, 104108 (1-8).

Submitted: 05/02/2022
Revised: 15/04/2022
Accepted: 25/04/2022
DOI: 10.48129/kjs.18531