

## RESEARCH ARTICLE

# goldenAGER: A Personalized Feature Fusion Activity Recognition Model for Elderly

DIANA NAGPAL<sup>1,2</sup>, SHIKHA GUPTA<sup>1</sup>, DEEPAK KUMAR<sup>3</sup>, ZOLTÁN ILLÉS<sup>4</sup>,  
CHAMAN VERMA<sup>4</sup>, AND BARNALI DEY<sup>5</sup>, (Member, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, University Institute of Engineering, Chandigarh University, Mohali, Punjab 140413, India

<sup>2</sup>Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Ludhiana, Punjab 141006, India

<sup>3</sup>Apex Institute of Technology, Chandigarh University, Mohali, Punjab 140413, India

<sup>4</sup>Department of Media and Educational Informatics, Faculty of Informatics, Eötvös Loránd University, 1053 Budapest, Hungary

<sup>5</sup>Department of Information Technology, Sikkim Manipal Institute of Technology, Majitar, Sikkim 737136, India

Corresponding authors: Zoltán Illés (illes@inf.elte.hu) and Chaman Verma (chaman@inf.elte.hu)

The work of Zoltán Illés and Chaman Verma was supported by the Department of Media and Educational Informatics, Faculty of Informatics, Budapest, Hungary.

**ABSTRACT** Physical and mental health are impacted as a person grows old. A Human Activity Recognition (HAR) system, which tracks a person's activity patterns and intervenes in case of an abnormal activity, could help elderly individuals to live independently. However, because of the strong intra class correlation between different activities, it is a challenging task to recognise such activities. Therefore, we proposed a personalized feature fusion algorithm, goldenAGER, which can be used to build as a model for abnormal activity recognition. In the initial stage, it extracts handcrafted HOG features and self-learned VGG-16 features to provide a rich description about the internal information of images. Then, the extracted features are provided as two different inputs to the deep neural network which are finally concatenated to classify the action type. The dataset is collected from the elderly volunteers over the age of 60 in a homogeneous environment consisting 10 classes of activities. The fusion of the features has resulted in 95% accuracy on primary dataset. The performance of the proposed model has also been tested on Microsoft Research (MSR) Action dataset giving accuracy of 93.08%. A comparison of our proposed model with the other existing models is also performed which shows that our model outperformed the existing models.

**INDEX TERMS** Abnormal behaviour recognition, deep neural network, elderly, feature fusion, human activity recognition, HAR.

## I. INTRODUCTION

Interpreting a human's movements in a video, also known as human activity recognition, is used to automatically deduce the activities and context from the video data by extracting detailed information about the human movements and the environment in which the activities are being performed. In a situation where the video is divided into segments such that each segment contains a human action, and the aim of the system is to successfully classify the activity into the category it belongs to [1]. Several HAR studies utilizing deep learning and machine learning have been carried out recently, but few

of them focus on developing a framework for elderly people. In 2017, there were 962 million persons in the world who were 60 or older, more than double the 382 million who were in that age group in 1980. Over 2.1 billion senior individuals are predicted to exist by 2050, a more than doubling of the current amount [2]. As a result of the enormous scientific advancement that has propelled human society into the global aging time over the past few decades, elderly care is the most recent use of HAR. Our ability to live longer and better lives has changed as a result of this development. With these developments, there is rising optimism and hope that most senior citizens would be able to live independently in their homes and take part in communal life. Various security methodologies can be used to provide security to the individuals [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeeb Dey<sup>1</sup>.

Even while technology advancements have made it possible for the majority of senior citizens to lead healthy and active lives, it is still dangerous for elderly individuals to live alone as they age because of both physiological issues and loneliness. Artificial intelligence is becoming more popular for HAR due to its capacity for self-improvement and reliable classification models [4]. The challenge of deriving useful features from sensor data has become easier since deep learning was introduced to the HAR domain [5]. The advancement of deep learning models facilitates the creation of numerous models like CNN [6], [7], Transfer learning which allows for the reuse of previously trained information by training a recognition model on a set of data and then using that model with a different testing dataset such as Inception [8], [9], ResNets [10], VGG-16 [11] and hybrid deep learning model development such as fusion of CNN-LSTM [12], Inception and ResNets [13].

Unfortunately, all of these algorithms have not been used effectively on elderly dataset. Since elder persons movements are different from young people because of various reasons majorly health issues, the model designed particularly for them cannot be tested on young people dataset as the results might differ in real-time. An AI based framework has been proposed for managing these kind of risks [14], [15]. A system whose objective is to recognize the elderly activities must have efficient prediction capabilities and high recognition rate. These traits can be used to predict the abnormal behavior of an elder person living alone. Hence the system will help to prevent any kind of mishappening in advance by raising an alarm to the caretaker.

The proposed research work will be able to answer the following research question: Research Question: How to propose a new workflow to address the HAR problem for elderly and evaluate it on real-time dataset?

To answer this question, we have proposed a personalized multi-input feature fusion method for elderly (goldenAGER), which would be able to extract light, illumination and shadowing invariant features using HOG and complex patterns using VGG-16 to recognize similar activities. To overcome the difficulty of the high dimensionality of the features, these feature maps are passed through a deep neural network. The final feature vector is then concatenated to recognize different actions. The proposed approach will aid in the development of strong HAR system to classify and recognize elderly activities in a more accurate way. There is no publicly available dataset available till date that provides videos data for elderly activities. Therefore, we have made a HAR dataset, in which the human actions data are collected from elderly people. The videos of the elderly volunteers were captured using a static camera. The proposed approach has been evaluated on primary dataset and publicly available Microsoft research (MSR) action detection dataset. Our model performs well on this dataset. The major contributions of this work can be summed up as follows:

- 1) We proposed personalized multi-input feature fusion model which takes into account both handcrafted and

self-learned features as two different input sources to classify the actions performed by elderly.

- 2) A novel algorithm, goldenAGER, is proposed to overcome the difficulty of extracting essential features from videos in order to recognize the activities of an elderly person.
- 3) During experiments, the proposed model is tested against our primary dataset and a standard MSR action detection dataset, yielding high recognition performance of 95 and 93.08 percent, respectively. Our model outperforms other approaches, demonstrating its suitability and viability for use in applications for elderly monitoring and abnormal activity prediction.

The paper's organizational structure is as follows: Section II focuses on the related study of HAR approaches using machine learning and deep learning algorithms. Section III elaborates the process of primary data collection and description of publicly available dataset used in the present research. Section IV explains the feature extraction process in detail. The design of algorithm, goldenAGER and the model is presented in Section V. Section VI presents experimental results and evaluates the algorithm on extensive performance metrics. This section also includes discussion on the performance of proposed algorithm in comparison to the existing algorithms. Section VII concludes the real crux of the proposed methodology, also including the future scope.

## II. RELATED STUDY

Human activities are the various behaviors that a person can perform over time in a certain setting. There is no perfect mechanism for recognizing human activity, therefore, academicians have been working on this subject for more than 20 years. Because it is easily influenced by numerous elements, including background noise, varying lighting, illumination, and intra-class variability, human activity is too complex to be recognized [9], [16], [17].

### A. TRADITIONAL MACHINE LEARNING-BASED APPROACHES FOR RECOGNISING HUMAN ACTIONS

A review and description of traditional machine learning methods using decision trees, k-nearest Neighbours (KNN), and support vector machine (SVM) was published by Jobanputra et al. [18]. Based on the dataset utilised, the model trained and the best features extracted, the system has been tested and accuracy has been validated. This study reveals that there are numerous difficulties with the conventional machine learning methodologies. For instance, while collecting data from sensors, it is mandatory for a person to wear many sensors, and the location of these sensors becomes a significant problem. Even when using typical handcrafted machine learning approaches, feature extraction can be challenging.

To evaluate the efficacy of human actions recognition using classic classification algorithms including SVM, KNN, random forest (RF), and Naive Bayes, Subasi et al. [19]

conducted a comparative analysis. Comparing their outcomes reveals that SVM has the highest classification accuracy. By monitoring various activities, such as walking and running, sitting and standing, Kececi et al. [20] investigated machine learning algorithms for user authentication. By implementing ten different machine learning algorithms on the HugaDB database, they proposed a person recognition system. The technique provides significant accuracy, and samples from 18 subjects were used to collect the data.

Zhang et al. proposed enhancements to the best-in-class activity recognition system, particularly for the activity representation and classification methods [21]. In this analysis, representation and classification methodologies are reviewed and examined. They describe current writing using a dataset they have created as well as point-by-point scientific categorization methodologies.

Traditional machine learning techniques have shown outstanding results for local feature extraction during the past ten years, according to Wei et al. [22]. Traditional and hand-crafted algorithms still have a number of drawbacks, like the fact that they take a long time to complete and that feature engineering selection is a highly challenging task because of high-dimensional features and sophisticated processing.

Now-a-days, researchers are using deep learning to create new methods for improved video-based human activity recognition because of the shortcomings and difficulties in standard machine learning approaches.

## **B. DEEP LEARNING BASED APPROACHES FOR RECOGNISING HUMAN ACTIONS**

Deep learning-based approaches employ the current end-to-end architecture to learn and represent high-level features while also categorizing them, in contrast to standard machine learning-based methods. These end-to-end designs learn the ideal qualities and adjust the parameters based on the data [23].

Liu et al. [24] suggested a technique for identifying humans in a brief video and spotting single-human movement that also involves human-human conversation. According to Hammerla et al. [25], deep, convolutional, and recurrent algorithms have been used that outperform state-of-the-art methods. They have also given directions for researchers who want to apply deep learning to address challenges.

By choosing to focus on specific aspects in the input frame, Muhammad et al. [23] introduced a bi-directional long-short-term memory-based attention method for detecting human actions from movies. For the extraction of the spatial and temporal patterns in different human actions from movies, He et al. [26] proposed a DB-LSTM system. Although deep learning has achieved notable success recently, action recognition remains a difficult issue due to occlusions, crowded backgrounds, camera motions, etc.

Guo et al. [27] integrated Wi-Fi signals along with video streams, employed convolution neural networks, and applied statistics analysis approaches to improve the recognition of

human actions in movies. Nieto-Hidalgo et al. [28] suggested a vision-based gait analysis technique that may be applied to a Smartphone with cloud computing. Cui et al. [29] created the YOLO-NSG (you only look once) method in the area of deep machine learning to lighten the load of inspectors by automatically spotting crime scenes in a video surveillance.

A capsule neural network was suggested by Basu et al. [30] to perceive human action coupled with visual layout. This is accomplished using deep CNNs. A new architecture based on 2D CNN was proposed by Gholamrezai and Taghi Almodarresi [31] to recognize human actions using only convolutional layers. The performance of the model will not change when the pooling layers are removed, but the computing time will significantly drop. HAR has developed into a significant topic of research for computer vision over the past ten years. Due to its superior performance, deep learning has been the primary focus of most experiments. But, extracting the essential features from vision-based data is still a challenging task [32]. Multi-feature extraction methods are used to extract the complex patterns in an image [33].

## **C. MULTI-FEATURE FUSION BASED APPROACHES TO RECOGNISE HUMAN ACTIONS**

Kim et al. [34] recognized nine older people's everyday routine activities by extracting skeleton joint features. They suggested a video based HAR framework that identifies typical indoor activities of elderly persons based on their skeleton's joint characteristics. Each human silhouette is given a collection of 23 joints after depth maps are processed to track them and create skeletal data on body joints. Following that, skeletal joints features are generated using the joint information. Finally, a hidden Markov model is employed to identify a variety of human behaviors utilizing these characteristics. The proposed approach outperformed as compared to the other existing approaches. Kaur and Gupta [35], [36] presents major challenges for the development of AI based system.

In order to categorize actions like walking, talking, etc., Seemanthini et al. [37] introduced a novel technique based on cluster segmentation for recognizing human action. The proposed model used HOG classifier for feature extraction. In contrast to current approaches, this system is able to accurately identify human objects of various sizes and shapes, producing a proficient action recognition method.

Khan et al. [35] integrated a hybrid model for activity recognition which makes use of a convolutional neural network (CNN) for extracting spatial properties and a long short-term memory (LSTM) network for learning temporal information. The accuracy achieved was better than existing approaches using the proposed approach. They have also proposed a new dataset with 12 number of classes of human actions.

Hayat et al. [2] implemented existing approaches to identify elderly activity. They applied a publicly available University of California Irvine (UCI) dataset to identify human activity. The study concludes that out of all machine learning

classifiers, SVM performs better but with a very low computation time, and out of all other available deep learning models, LSTM performs better.

Zerrouki et al. [38] developed an effective human action recognition technique that benefits from the adaptive boosting algorithm's excellent discrimination ability. Using experimental data from two publicly accessible fall detection datasets from the Universidad de Málaga and the University of Rzeszow, they evaluated the efficacy of their methodology. Furthermore, they have shown that when it comes to distinguishing different human gestures, the suggested method beats state-of-the-art classifiers based on neural networks, K-nearest neighbor, support vector machines, and naive Bayes. A new approach to expressing and detecting human activity was proposed by Hbali et al. [39]. The proposed approach computes spatial and temporal channels from 3D skeleton coordinates and combines them to represent each frame of a human activity sequence. The technique employing overly randomized trees was trained and validated using the MSR Daily Activity 3D and MSR Action 3D datasets.

Huan et al. [40] provided a method for extracting multi-layer features from hybrid CNN and BiLSTM. To improve the performance of human complex activity recognition, features from different sets that are obtained via trained networks, manual extraction, and location information are combined to produce mixed features. The proposed approach is tested on the PAMAP2 and UT-Data datasets. The results demonstrate a method that greatly outperforms the state-of-the-art methods now in use. This method is based on a number of qualities addressed in this research.

By Kiran et al. [41], a multi-feature fusion model based on HAR was suggested. The global average pool and fully connected layers are used to extract features. The canonical correlation analysis combines the characteristics of the two levels. The chosen characteristics are passed to several classifiers in the end for final classification. For the final classification, the selected attributes are sent to many classifiers which performed quite well to classify the actions.

Three networks are used in Chan et al. [42]'s unique approach to determine the position of the human and other objects in the scene that are the most related to the human, and the entire situation, which includes the actors and everything else in the scene. Three parallel networks are utilized to incorporate these factors, followed by feature fusion and convolutional neural networks. An SVM classifier then performs the classification.

A spatio-temporal feature fusion approach was put forth by Yao et al. [43] to gather the crucial visual data for recognition. Moreover, they expanded on the bag of words representation by adding a compressed spatio-temporal video representation. Their tests on two well-known datasets demonstrate effective performance.

Huynh-The et al. [44] proposed a feature fusion approach called deep convolution neural network (DCNN) to simultaneously extract spatiotemporal information. In this method, the SVM classifier is used to train the hybrid feature, which

mixes the deep and handmade features. The model was tested on three benchmark datasets for human activity recognition.

Table 1 shows summary of the existing approaches. A multi-modal learning methodology utilising video and wi-fi feature fusion was developed by Guo et al. in [27]. Using proposed Vi-Wi15 dataset, they have achieved 94.2% accuracy. Basu et al., [30] proposed capsule neural network by extracting features using VGG-16 to recognize indoor home scenes. The model was trained and tested on public dataset with accuracy 76.9%. Gholamrezai and Taghi Almodarresi et al., [31] proposed a 2DCNN model with no pooling layer for activity recognition. They have used publicly available HAR dataset and achieved 95% accuracy. Kim et al., [34] proposed skeleton joints feature detector by concatenating the features collected from centroid point, magnitude and joints distance. They have created an elderly dataset with 9 activities involving eating, drinking, lying down, stand up, walking, exercise etc. Using the proposed method, they have achieved 84.33% accuracy. Seemanthini and Manjunath [37] proposed a model based on cluster segmentation. They used HOG+SVM to recognize the human activity. A publicly available dataset is being used in this paper which gives 89.59% recognition accuracy. Number of classes and activity types in the dataset are not mentioned in the paper. Since this model is trained and tested on publicly available dataset, the results might vary if this model will be used on elderly people. Hbali et al., [39] proposed a skeleton-based approach for elderly monitoring systems. They have used publicly available datasets involving pickup, throw, two hand wave, hand clap etc. activities with overall accuracy of 86.1%. Similarly, Abdelbaky and Saleh Aly [45] proposed spatio-temporal feature fusion HAR model. The model gives 93.3% accuracy but, the model has been trained and tested on sports dataset. All the activities considered for recognition are also sports activities such as dive, golf, kicking, Ride Horse and Run etc. Therefore, this model is also not applicable for elderly people. Khaire et al., [46] proposed decision level fusion approach by fusing the ConvNet and SoftMax scores. They have used 4 publicly available datasets with different no. of classes including activities swipe, wave, clap, throw, boxing, squat, draw circle etc. The proposed model achieved 94.60%, 92.82%, 93.06%, 96.26% accuracy respectively on each dataset. But because these datasets are collected from normal people, result may vary on elderly people. Zhang et al., [47] proposed handcrafted and learned feature fusion method in which the features collected from discrete wavelet techniques and CNN-RNN model are fused with accuracy around 95.1%, 70%, 72.5% respectively on three different publicly available datasets. But these datasets have been collected from YouTube videos for people of all ages not specifically elders, therefore, the results might vary for elderly people. Most of the research has been focused on healthy young people, therefore, these approaches are partially applicable to be used as a framework for elderly monitoring and abnormal activity recognition systems. Our proposed method used handcrafted and learned feature fusion

TABLE 1. Summary of the existing methods.

Reference No.	Feature Extraction Method	Dataset	No. of classes	of Activity Type	Accuracy	Applicability
[27]	Video and Wi-Fi feature fusion	Primary	15	Basic human activities	94.20%	Partially applicable because it is trained and tested on Primary dataset but not specifically for elderly, therefore results might vary for elderly.
[30]	VGG-16	Public	5	Indoor Scenes	Home 76.9%	Partially applicable because it is trained and tested on Publicly available dataset for young people, therefore results may vary for elderly.
[31]	2DCNN	Public	6	Indoor Scenes	Home 95%	Partially applicable because it is trained and tested on Publicly available dataset for people of age between 19-48 years, therefore results may vary for elderly.
[34]	Skeleton Joints feature extractor	Primary	9	Daily Routine activities	84.33%	Fully applicable because it is trained and tested on primary elderly dataset, therefore can be used as a framework to predict abnormal elderly activities but recognition accuracy is less.
[37]	HOG+SVM	Public	-	-	89.59%	Partially applicable because it is trained and tested on Publicly available dataset for young people, therefore results may vary for elderly.
[38]	Five Occupancy Zone Generation	Public	6	Daily Routine activities	93.91%	Partially applicable because it is trained and tested on publically available elderly dataset, therefore can be used as a framework to predict abnormal elderly activities.
[39]	Thresholding Method with inverse HWT	Public	7	Daily Routine activities	86.1%	Partially applicable because it is trained and tested on primary elderly dataset, therefore can be used as a framework to predict abnormal elderly activities.
[42]	ResNet18	Public	10	Daily Routine activities	87.13%	Partially applicable because it is trained and tested on Publicly available dataset on normal people, therefore results may vary for elderly.
[43]	Spatio-temporal feature fusion	Public	6 classes and 11 classes respectively	Sports Activities	74.38%	Not applicable because it is trained and tested on Publicly available sports action dataset on young healthy people, which may not be applicable for elderly.
[45]	Spatio-temporal feature fusion	Public	10	Sports Activities	93.3%	Not applicable because it is trained and tested on Publicly available sports action dataset on young healthy people, which may not be applicable for elderly.
[46]	Decision Level Fusion	4 Public Datasets	27 classes, 20 classes, 12 classes, 8 classes respectively	Daily routine Activities	94.60%, 92.82%, 93.06%, 96.26% respectively	Partially applicable because it is trained and tested on Publicly available dataset on normal people, therefore results may vary for elderly.
[47]	Handcrafted and Learned feature fusion	3 Public Datasets	11 classes, 51 classes, 6 classes respectively	Daily routine Activities	95.1%, 70%, 72.5% respectively	Partially applicable because it is trained and tested on datasets collected from YouTube videos, not specifically elders, therefore results might vary for elderly.
<b>Proposed approach</b>	<b>Handcrafted and Learned feature fusion</b>	<b>Primary and MSR Action Dataset</b>	<b>10 classes, 10 classes respectively</b>	<b>Daily routine Activities</b>	<b>95%, 93.08% respectively</b>	<b>Fully applicable because it is trained and tested on primary dataset specifically elderly and can be used as a framework to predict abnormal elderly activities</b>



FIGURE 1. Sample images of different activities performed by elderly.

method with accuracy 95% and 93.08% respectively. Since our model has been trained and tested on dataset collected specifically for elderly, it is most suitable to recognize elderly people activities.

As discussed in the previous section, traditional hand-crafted machine learning algorithms and deep learning algorithms have shown remarkable results for feature extraction and human activity recognition, but due to some limitations and challenges in extracting features using these methods, multi-feature fusion methods are used.

### III. OBJECTIVES

After a thorough analysis on the existing approaches with a focus on present limitations, the following objectives have been set forth to address the significant research gap.

**Objective 1:** To propose a HAR model that can be incorporated as an application to continuously monitor the activities of an elderly person and predict abnormal activity.

**Objective 2:** To develop a novel algorithm which recognizes the elderly activities such as Carry, Clap Hands, Pickup, Pull, Push, Sit Down, Stand Up, Throw, Walk and Wave Hands.

**Objective 3:** To explore the significant features such as important landmarks, edges, textures and complex patterns in the videos to identify these activities.

**Objective 4:** To evaluate the goldenAGER algorithm on extensive performance metrics such as Precision, Recall, F1-score, Cohen's kappa Coefficient and ROC curve.

**Objective 5:** To test and compare the proposed algorithm with existing algorithms.

### IV. DATA COLLECTION

The dataset used in this work has been collected from elderly volunteers over the age of 60 which contains video sequences of 10 classes of continuous human actions. The sequences were taken over homogeneous environment using a static camera.

The sample images for each of the 10 kinds of elderly movements are shown in Figure 1. Videos of 10 sec were captured for each class of activity which are then converted into image frames to provide as an input to feature extractor. 30 frames were displayed every second. There were around 2% of the frames in which same information was recorded with very slight difference. Therefore, we have kept only one frame. The final frame rate was 15 frames per second. The covered activities are carry, clap hands, pickup, pull, push, stand up, sit down, throw, walk and wave hands. Sample

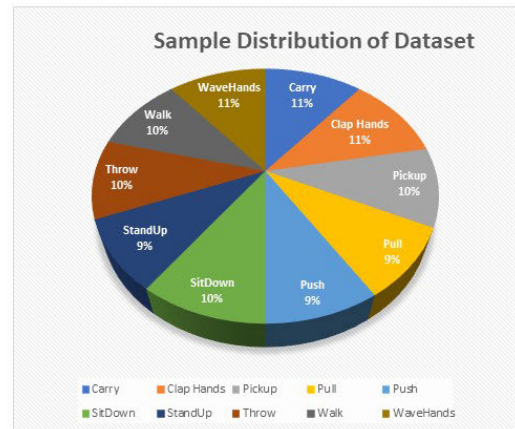


FIGURE 2. Sample distribution.

distribution of each class of activity is shown in Figure 2. Out of 20270 total images, the percentage of samples in each activity are represented.

The quantity of samples that were gathered for each activity individually is de-tailed in Table 2. The data is split into training and testing data with an 80/20 split, meaning that To determine how the algorithm will act in the presence of unobserved data, 20% of the samples are used for testing and 80% of the samples are used for training.

MSR Action Detection Dataset: We have also used MSR Action dataset to validate our model. For experiments on action detection, this dataset is used. It has ten subjects and twenty different action kinds. Each subject carries out the tasks twice or three times. This dataset has the advantage of using action sequences from several angles. For instance, there are significant differences between several ways that the same action is implemented [48]. We have compared the performance of our model on MSR Action dataset also. The frames were captured from video sequences as shown in Figure 3.

### V. FEATURE EXTRACTION

#### A. EXTRACTING HOG FEATURE DESCRIPTORS

Since humans can appear in a variety of ways and adopt a variety of poses, detecting them in images can be difficult. The first need is a strong feature set that enables clean human form discrimination even in cluttered backgrounds with challenging lighting. To fulfil this requirement, [49] proposed HOG for human detection.

TABLE 2. Samples collected for each activity, in number.

Activity	Training Data	Testing Data
	Number of Samples	Number of Samples
Carry	1784	446
Clap Hands	1784	446
Pickup	1622	405
Pull	1459	365
Push	1459	365
Sit down	1622	405
Stand up	1459	365
Throw	1622	405
Walk	1622	405
Wave Hands	1784	446
<b>Total</b>	16217	4053



FIGURE 3. Sample images of different activities performed by elderly.

This section talks about how to extract HOG features of humans. The steps below demonstrate the entire method for obtaining HOG’s characteristics:

- 1) Pre-processing: RGB frames that have been collected from the dataset are of size 480\*640. All the activities have been resized to the new size using Python openCV libraries.
- 2) Calculate the Gradients of the image: Combining the image’s magnitude and angle produces the gradient.  $G_x$  and  $G_y$  are initially calculated for each pixel in a block of  $3 \times 3$  pixels. The following formulas are used to calculate  $G_x$  and  $G_y$  first for each pixel value:

$$G_x(r, c) = I(r, c + 1) - I(r, c - 1) \tag{1}$$

$$G_y = I(r - 1, c) - I(r + 1, c) \tag{2}$$

where r and c are rows and columns respectively. After that, each pixel’s Magnitude and Gradient are calculated from the following formula:

$$Magnitude(\mu) = \sqrt{(G_x)^2 + (G_y)^2} \tag{3}$$

$$Angle(\theta) = \tan^{-1} G_x/G_y \tag{4}$$

Figure 4 explains the process to calculate gradient magnitude and direction in detail.

- 3) Block Normalization: Two cells ( $8 \times 8$ ) are combined to make a  $16 \times 16$  block. This clubbing is carried out in an overlapping way with an 8-pixel stride by concatenating each cell’s nine-point histograms to create a 36-point vector for all four cells in the block., represented as

$$fb_i = b_1, b_2, b_3, \dots, b_{36} \tag{5}$$

where  $fb_i$  is one feature vector belonging to one block. Then values of  $fb_i$  are normalized by L2 normalization. By normalising the blocks, the impact of differences in contrast between images of the same class is lessened.

- 4) Getting HOG feature vector: The image has been divided into  $8*8$  cells, the cells per block are (2, 2).  $16*16$  sized block has been taken and the final feature vector has been calculated using all the 126 features that are combined together. All the features have size 36, so  $126*36$  i.e., 4536 feature vector has been formed.

$$F = f_1, f_2, f_3 \dots, f_k \text{ where } k = 4536 \tag{6}$$

$f_1, f_2, f_3 \dots, f_k$  are the feature dimensions.

Figure 5 shows the visualization of HOG features on a sample image from dataset.

**B. EXTRACTING FEATURES USING VGG-16**

An extremely deep neural network called VGG-16, developed by [50], is used for both feature extraction and image classification. We have used VGG-16 in this work as a feature extractor. The 14 million images in the ImageNet dataset, which is divided into 1000 classes, were used to pretrain the VGG-16 model. The ImageNet collection contains images with RGB channels and a fixed size of  $224*224$ . As a result, our input is a matrix of (224, 224, 3).

As seen in Figure 6, this model analyses the input image and generates a 1000-value feature vector. Instead of training with a smaller dataset, it is a good idea to import the VGG-16 weights. The Visual Geometry Group, University of Oxford’s data is used to import the weights for the VGG-16 model [50]. The weights from the source are provided in MATLAB format. The provided weights are compatible with MATLAB, and MatConvNet is a MATLAB toolbox for CNN. The weights of the VGG-16 are converted in this study

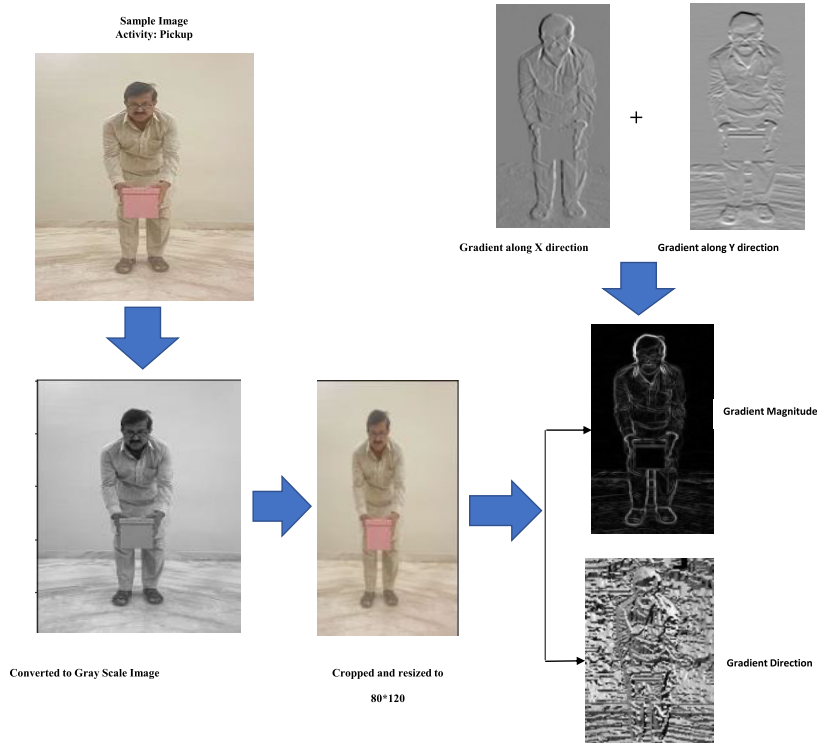


FIGURE 4. Process to calculate gradient magnitude and direction.

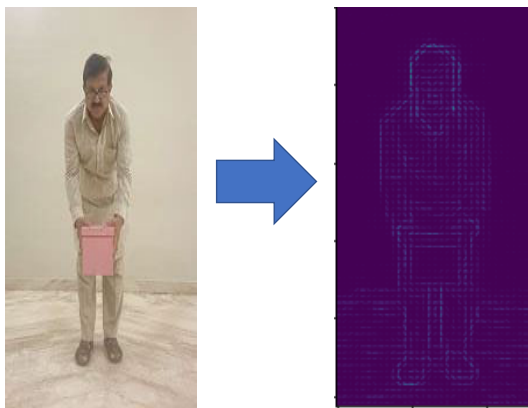


FIGURE 5. Visualization of HOG features.

from MATLAB format to a Keras-compatible format using machine learning libraries. The total trainable parameters are 13.8 million in VGG-16 after going through each layer. Transfer learning helps the researchers to focus on the real problem while saving computation time. In the proposed model, 1000 deep features from the imported model are combined with the HOG features. Figure 7,8,9, 10 represents VGG-16 low-level and high-level features in detail.

### C. INTRA-CLASS VARIATIONS

The same person repeating the same activity with the exact same execution is extremely uncommon. Moreover, even

when executing the same action, every person behaves differently. For instance, walking action may be different for every person as each individual walks at his own pace, throwing action may be different and so on [32]. After concatenating the features, intraclass variation of different action classes is analyzed by finding spearman’s correlation matrix [51], [52] as shown in equation (7).

$$Corr(\phi_i) = 1 - 6d^2(x, y)/n(n - 1) \quad (7)$$

where  $\phi_i$  is the correlation coefficient and n is the total feature samples. where  $\phi_i$  is the correlation coefficient and n is the total feature samples.

$$d(x, y) = \sum |m(x) - n(x)| \quad (8)$$

The distance between two feature samples for the same class in the dataset, x and y, is measured by  $d(x,y)$ . If the distance value is minimal, the characteristics are identical. An ideal action recognition system should be able to generalize over variations of a class in order to recognize a previously unidentified instance. The spearman’s correlation coefficient’s values for the 10 classes in the primary dataset are Carry (+0.41), Clap Hands (+0.59), Pickup (−0.42), Pull (−0.08), Push (−0.21), Sit Down (+0.77), Stand Up (+0.89), Throw (−0.91), Walk (−0.82) and Wave Hands (+0.73). The correlation values vary between -1 and +1. Negative values shows that the feature samples are not correlated or weakly correlated to each other which means that there is a large variation between the different feature samples for the same



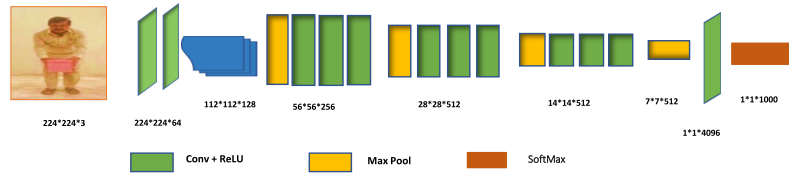


FIGURE 6. VGG-16 Architecture for feature extraction at abstract level.

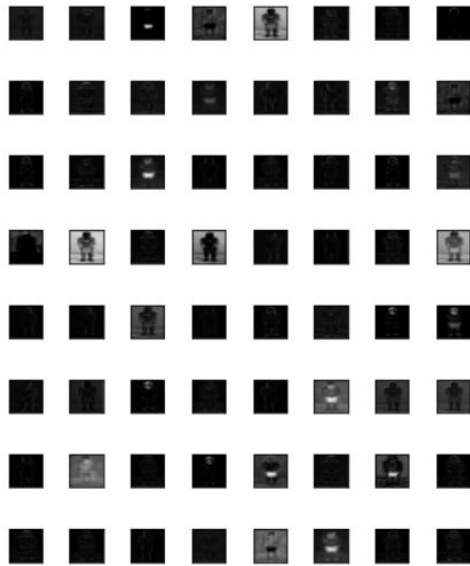


FIGURE 7. Extracting lower-level features (minor details).

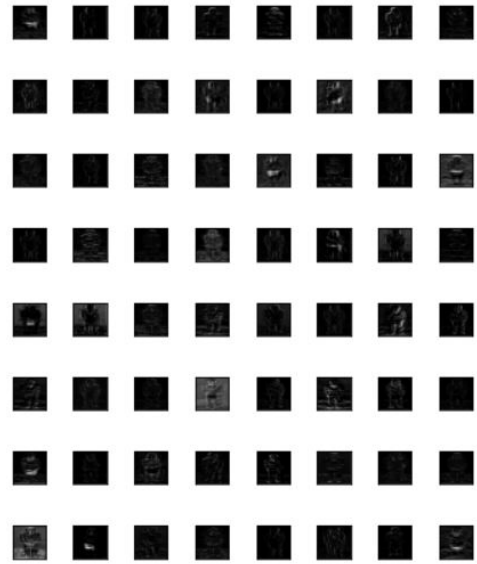


FIGURE 9. Extracting lower-mid features (important landmark points).

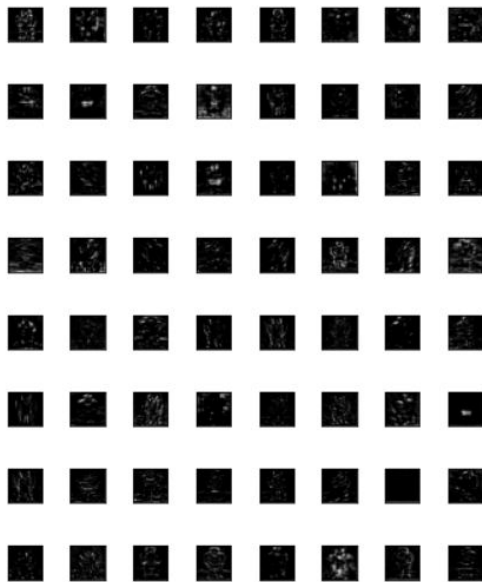


FIGURE 8. Extracting higher-mid-level features (textures and patterns).

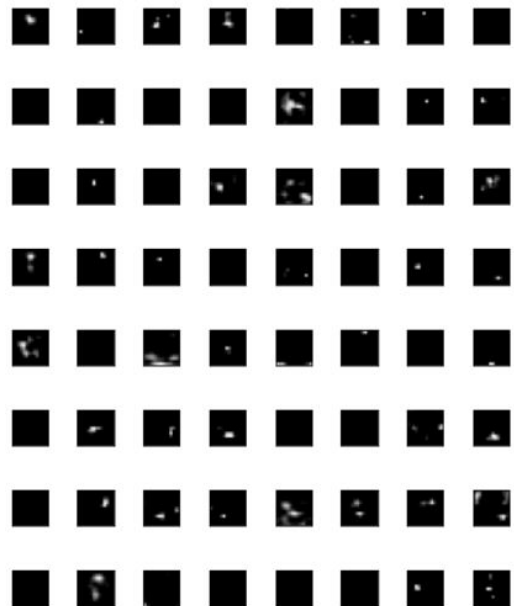


FIGURE 10. Extracting high level features (complex patterns).

class. It can be seen from the values that Pick Up, Push, Pull, Throw and Walk classes have large intra class variation because of the directional changes. Moreover, one person can

use both hands for pick up, push and pull actions while the other can use one hand.

An action recognition model is proposed to handle intra-class variations [53] using HOG+HOF descriptors with accuracy 93.13%. Our proposed method outperforms the existing method. The block normalization step used in HOG feature extraction process improved the performance of HAR model by extracting the discriminative features.

Using a fine-tuned CNN model enhances generalization rather than training the model on smaller datasets. A transfer learning strategy called “fine-tuning” focuses on using the information gained from addressing one problem to solve another that is comparable. By using pre-trained weights, which are created from larger databases through fine-tuning, the overfitting issue can also be avoided. This happens because of the multiple layers and characteristics that CNNs have. The CNN models usually identify patterns that are especially pertinent to the smaller datasets on which they are trained. When the model uses fine tuning, the filters can be reused. VGG-16 is used in this paper as a transfer learning strategy which has a significant capacity for discrimination. As a result, the HOG and VGG-16 combination handles intraclass variation very effectively.

## VI. ALGORITHM AND MODEL DESIGN

Feature fusion makes it easier to properly learn features for the rich description of their internal information. After collecting features from HOG feature descriptor and VGG-16 feature extractor, the feature maps are passed through three hidden layers of CNN and a dropout layer to reduce the number of features. After That, 256 HOG features and 256 VGG-16 features are concatenated. The final feature vector is then passed through two input layers which contains features and labels separately which in turn classify the output actions.

### A. BENEFITS OF MULTI-FEATURE FUSION

- Using HOG feature descriptor, intensity gradients distribution and edge directions can be described to find local object shape and appearance. After the image is separated into cells, which are connected, small areas, and for the pixels within each cell, a gradient histogram is created. The combination of these histograms becomes the feature descriptor. The local histograms are contrast-normalized to increase accuracy by generating an intensity measure over a larger area of the image, referred to as a block, and using this value to normalize all cells within the block. Better invariance to variations in illumination and shadowing is the effect of this normalization.
- To discriminate between various human activities, features at low-level like edges and corners are required. Additionally, the features may become locally intensive to illumination conditions after normalizing the cells over wider areas, like blocks.
- Transfer learning is a machine learning technique that uncovers hidden patterns and important information in one particular domain before applying and transferring this knowledge to an unrelated but companion area.

TABLE 3. List of abbreviations and notations used in goldenAGER.

Symbol	Notation
$m_t$	Aggregate of gradients at time t
$\beta$	Decay Rate
$\delta L$	Derivative of the loss function
$\delta w_t$	Derivative of the weights at time t
$v_t$	Sum of square of gradients
m	Image size
f	Filter size
ch	Number of channels
p	Padding
s	Stride Number
n	Number of filters

In transfer learning, we take advantage of the model’s ability to generalize what it has learnt for one task to a different, related issue. As compared to building a model from scratch, transfer learning uses pre-trained models that require less computing work to fine-tune. VGG-16 is a pretrained model on the ImageNet dataset containing 14 million images belonging to 1000 number of classes. ImageNet dataset is exclusively trained for object recognition. We have used VGG-16 in this paper because of its efficiency on popular datasets such as Stanford-40 action dataset [14] and Multi-View Human Action Recognition (MVHAR) datasets [15].

- HOG and VGG-16 can efficiently operate together to collect crucial action features. Important low-level action features are gathered via the HOG features. The VGG-16 characteristics generalize the local features collected from human activity to deal with high-level features. VGG-16 extracts features at an abstract level by employing various action patterns. Given that HAR relies on two distinct features that are produced using two different approaches, the multi-feature fusion of HOG and VGG-16 improves the effectiveness of HAR.

The model is divided into two phases. In Phase 1, features are extracted using two different methods, handcrafted feature extraction and transfer learning feature extraction. Both the methods are explained in detail in the previous section. After getting both feature descriptors, the output of Phase 1 is passed as two different inputs in Phase 2. In this phase, goldenAGER, multi-input feature fusion algorithm is applied to classify various elderly activity classes to recognize their actions. Two feature sets are passed as two different inputs to the deep neural network. One input is the HOG feature vector and second input VGG-16 feature vector plus pre-trained VGG-16 weights. Along with the convolutional layers, activation functions are also used to ensure that only valuable information is gathered, speeding up computation. Pooling layers are used to aid with memory size issues so that the information acquired and its associated data are condensed to a smaller memory size. The model is explained in detail in Figure 12. List of abbreviations and notations used in goldenAGER are described in Table 3.

**Phase 1:** Feature extraction**INPUT:** Video sequences captured from static camera**OUTPUT:** Two Feature Vectors

- 1) Initialize Frame rate and image size= 480\*640 pi
- 2) Convert RGB images to Grayscale images
- 3) Resize the image to 80\*120 pi
- 4) **For** each pixel in a block of 3\*3 pixels **do**
  - a) Calculate the gradients  $G_x$  and  $G_y$
  - b)

$$G_x(r, c) = I(r, c + 1) - I(r, c - 1) \quad (9)$$

$$G_y = I(r - 1, c) - I(r + 1, c) \quad (10)$$

- c) Compute Magnitude and Direction of the Gradient

$$Magnitude(\mu) = \sqrt{(G_x)^2 + (G_y)^2} \quad (11)$$

$$Angle(\theta) = \tan^{-1} G_x / G_y \quad (12)$$

- d) Divide gradient image into 8\*8 cells
- e) Find histogram bins for each cell
- f) Group 2\*2 cells to form overlapping blocks
- g) Normalize the blocks.
- h) Get the HOG descriptor for the image

**End for**

- 5) Save HOG feature vector
- 6) Repeat step 1 for using VGG-16 feature extractor
- 7) Resize the image to 224\*224\*3
- 8) Use the pretrained weights of ImageNet dataset
- 9) Convert the image frames to NumPy array
- 10) Apply VGG-16 model
- 11) Save VGG-16 feature vector
- 12) **Return** HOG and VGG-16 feature vector

**Phase 2:** goldenAGER: multi-input feature fusion algorithm**INPUT 1:** HOG feature vector**INPUT 2:** VGG-16 feature vector + pre-trained weights**OUTPUT:** Recognized action class

- 1) Pass Input 1 to hidden layers
- 2) Evaluate individual neuron weights  $w_{ij}$  during training stages by passing training video data
- 3) **For** each Layer **do**
  - a) Apply ReLU activation function

$$F(x) = \text{Max}(0, x) \quad (13)$$

- b) Use ADAM optimizer to deal with non-stationary activities

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) [\delta L / \delta w_t] v_t \\ &= \beta_2 v_{t-1} + (1 - \beta_2) [\delta L / \delta w_t]^2 \end{aligned} \quad (14)$$

- c) Use Dropout Layer to prevent overfitting
- d) Apply Categorical Cross Entropy loss function to minimize error value

$$Loss = - \sum_{i=1}^{OutputSize} y_i \cdot \log \hat{y}_i \quad (15)$$

- e) Compute feature vector dimensions

$$\begin{aligned} &[m, m, ch] * [f, f, ch] \\ &= [(m + 2p - f) / s + 1] \\ &\quad * [(m + 2p - f) / s + 1], n \end{aligned} \quad (16)$$

**End for**

- 4) Repeat steps 1 and 2 for Input 2
- 5) Concatenate both the feature vectors
- 6) Pass concatenated feature vector to 2 different input layers, one for features and other one for labels.
- 7) Apply ReLU activation function to check the feature maps for any negative values
- 8) Use Dropout Layer again to prevent overfitting
- 9) Pass the final feature vector to trained model for classification
- 10) **Return** Recognized Action Class

Figure 11 depicts the proposed model architecture. Basically, the model takes two inputs. One from the HOG descriptor and the another one from VGG-16 feature extractor. CNN model is used to classify the actions. Deep neural networks that use convolutional layers, convolutional operations, activation functions, pooling layers, and SoftMax functions make up the CNN class. CNN can be directly fed an RGB image as an input. The RGB image is subsequently processed by the CNN model using a variety of filters, producing feature maps. The essential hyperparameters are size of the filter, total number of filters and padding type. The depth of the output feature map is determined by these hyperparameters. The size of the images after convolution can be determined using Equation 9.

$$(I * H)[m, n] = \sum_j \sum_k H[j, k] * I[m - j, n - k] \quad (17)$$

Another important hyperparameter is padding type. The two types of convolutions that are typically used are Valid Convolution and Same Convolution. Valid basically means there is no padding. Therefore, the size is reduced as a result of each convolution. Padding is used with Same Convolution to maintain the size of the matrix. The size of  $N \times N$  will be kept if we pad the image by  $(F - 1) / 2$  pixels on all edges.

The feature vector dimensions have been mentioned in Figure 11. After calculating the feature vector, ReLU activation function is applied to bring nonlinearity in the model. After that, by downsizing the image while keeping the essential features, we have applied pooling layer to reduce feature vector size. To extract the few most useful features from an input image, a block made up of a Convolution layer, ReLU, and max pooling layer is repeated. The 3-dimensional feature maps are then flattened to create 1-dimensional versions. The last fully connected layer has a SoftMax activation function that generates probabilities between 0 and 1 for the relevant output class. CNN model employed in this paper is VGG-16 shown in Figure 3.

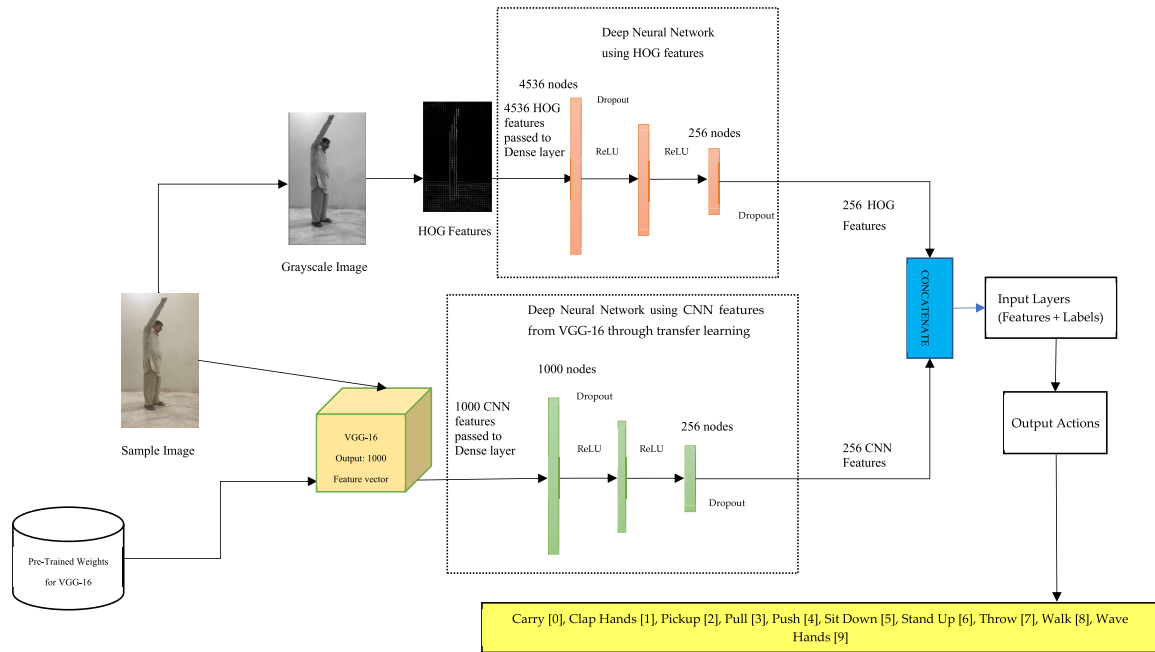


FIGURE 11. Proposed two-input model architecture for elderly activity recognition.

## VII. RESULTS AND DISCUSSIONS

The success of a HAR model depends upon the three major factors that are the model architecture, Loss function design and the Hyperparameters selection. In our proposed model, we have carefully worked on all these important factors.

### A. MODEL ARCHITECTURE

In our proposed model, HOG features and VGG-16 features are concatenated. The benefit of concatenation is that feature fusion provides rich description of the internal information about the activities being performed.

One of the main parameters to find HOG feature vector is cell size. If the cell size is small, it will not be able to classify between valuable and irrelevant features. On the other hand, if we choose a larger cell size, then it will compress the portion of the image and some important information may be lost. Table 4 shows the accuracy of HOG+SVM on primary dataset and MSR Action dataset. It can be seen from the table that A cell size of [32], [32] does not capture much shape features, whereas [4], [4] captures a lot of shape features but increases the HOG feature vector and slows down training. As a result, [16], [16] is the cell size used in this study. Another main parameter is the number of orientation bins. If we choose a smaller number of orientation bins, it will lead to information loss and if a large number is chosen, it will spread the information in all bins and lower the performance of HAR. A proper set of these two parameters makes the HOG feature extractor the most suitable one to categorize elderly actions peculiarities.

The benefits of using VGG-16 are that it is pre trained on millions of images which give more accurate description of

the features. Combining the two feature vectors will collect all the important information from the image. While HOG works on low level features, VGG-16 works on abstract level features to extract complex patterns from the image.

The preparation of HAR models necessitates a GPU. This model is trained on NVIDIA GeForce GTX 1650 with 16GB RAM. The model used Intel Core i5-10th Gen Processor with a clock speed of 2.50GHz. Keras has been used to build deep neural network architecture. We have used ReLU as an activation function because it brings non-linearity [46]. It will check the feature maps and if there are any negative values, it will replace them with zero. ReLU helps with making the model non-linear.

### B. LOSS FUNCTION DESIGN

To reduce the error, categorical cross entropy loss function is utilized. It is a loss function used in multi-class classification. The formula to calculate this loss function is as follows:

$$Loss = - \sum_{i=1}^{OutputSize} y_i \cdot \log(\hat{y}_i) \quad (18)$$

where  $\hat{y}_i$  is the  $i^{th}$  scalar value in the model output, Output size is the number of scalar values in the model output, and  $y_i$  is the equivalent target value.

The degree to which two discrete probability distributions can be separated from one another is well demonstrated by this loss. In this instance,  $y_i$  stands for the probability that event  $i$  will occur, and the sum of all  $y_i$  equals 1, indicating that only one event is possible. The minus sign ensures that the loss reduces as the distributions get closer to one another.

TABLE 4. HOG accuracy on primary dataset and MSR action dataset by changing parameters.

Number of orientation Bins	Block Size	Cells Per Block	Accuracy	No. of Features
9	(4,4)	(2,2)	82%, 83.42%	40716
9	(8,8)	(2,2)	85%, 84.69%	9576
9	(16,16)	(2,2)	85.92%, 86%	4536
9	(32,32)	(2,2)	88.06%, 89.57%	1944

TABLE 5. Details of the hyperparameters selection for the proposed model.

Feature Extraction Method/ Type of model	Hyperparameters	Values
HOG	Input Image	80*120
	Cell Size	8,8
	Number of Orientation Bins	9
VGG-16	Input Image	224*224*3
CNN Model	Input	Concatenated Feature Vector
	Number of Hidden Layers	03
	Activation Function	ReLU
	Number of Epochs	Adaptive
	Batch Size	20
	Dropout	20-50%
	Optimizer	ADAM
	Loss Function	Categorical Cross Entropy
	Output Classes	10 Output Classes Carry, Clap Hands, Pickup, Pull, Push, Sitting, Standing, Throw, Walking and Wave Hands

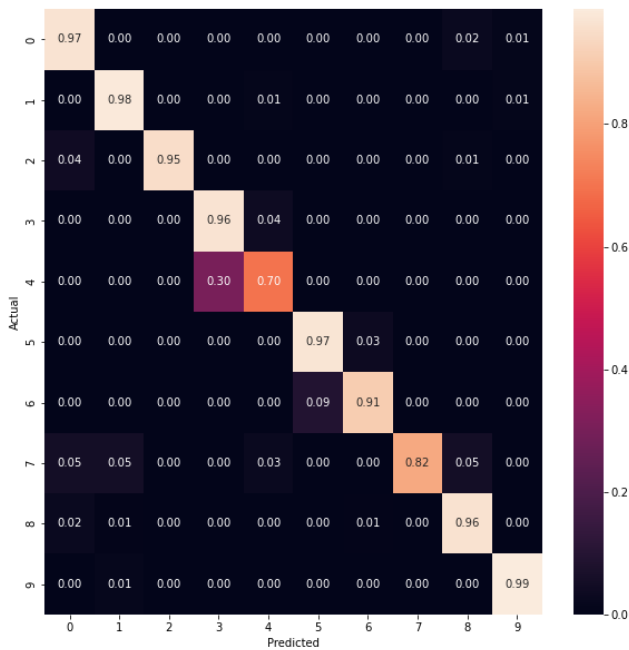


FIGURE 12. Confusion matrix representing the recognition accuracy using primary dataset).

C. HYPERPARAMETERS SELECTION

Hyperparameters selection is one of the important factors for the selection of a HAR model. Table 5 depicts the details of the hyperparameters selection in the proposed model. As discussed in the previous sections, cell size and no. of orientation bins used are essential hyperparameters for HOG. For VGG-

16 model, the output image size and output feature map dimensions are the important hyperparameters as shown in Figure 12. Number of hidden layers, epochs, dropout rate, Optimizer etc. are another important hyperparameters for CNN model.

The model was trained and tested on real-time elderly dataset showing which shows outstanding performance of the model. Figure 12 presented the confusion matrix plotted for each class for primary dataset and Figure 13 depicts the confusion matrix for MSR action detection dataset. The model is getting a little bit confused between push and pull because of strong intra-class correlation between these two activities. For all the other activities, the model is behaving well. It is showing 99% accuracy in case of wave hands, 96% accuracy in case of carry and sit down, 96% for push and walk, 95% for pickup activity, 91% for stand up and 82% for throw. Figure 14 and Figure 15 represents training and validation accuracy of the proposed model on two different datasets.

D. PERFORMANCE MEASURES

The precise classification of each class is equally significant in the perspective of F-measure (F1). The model can be evaluated more effectively than the precision because when computing the score, it takes into account each class’s recall and precision. Following are the formulas to calculate precision, recall and F1 score.

Precision:

$$P = \frac{TP}{TP + RP} \tag{19}$$

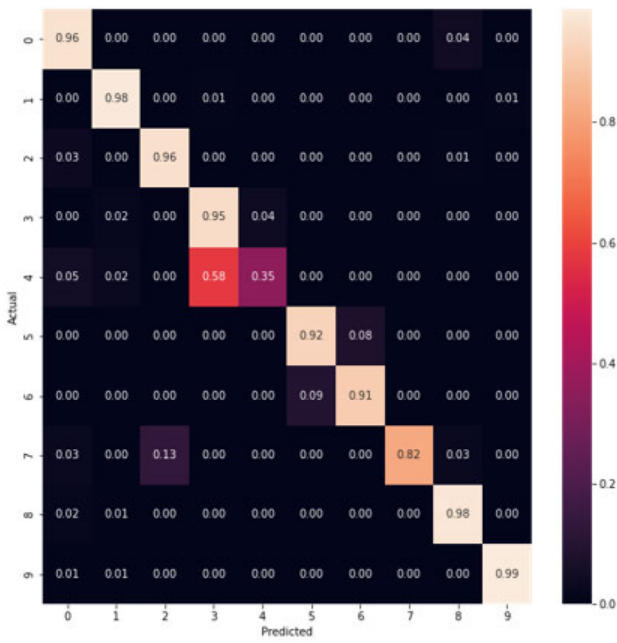


FIGURE 13. Confusion matrix representing the recognition accuracy using MSR action detection.

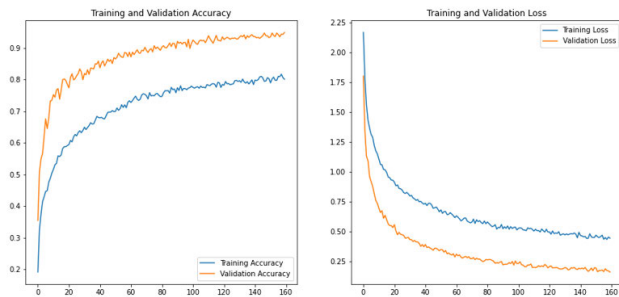


FIGURE 14. Training and validation accuracy of the proposed model on primary dataset.

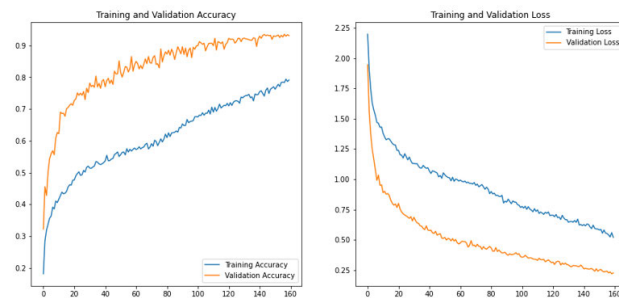


FIGURE 15. Training and validation accuracy of the proposed model on MSR action detection dataset [39].

Recall:

$$R = \frac{TP}{TP + FN} \tag{20}$$

F1-Score:

$$F1 = \sum_i 2 * w_i \frac{Precision_i * recall_i}{Precision_i + recall_i} \tag{21}$$

where TP and FP stand for the corresponding true and false positive counts, and FN stands for the corresponding true negative count. Class weighting is used to address class inequality. If  $w_i = n_i/N$  is the percentage of samples from the  $i$ th class,  $n_i$  is the total number of samples according to their sample proportion, and  $N$  is the number of overall samples. Table 6 depicts the precision, recall and F1-score for each class. It can be seen from the table that precision is greater than 0.5 in all the cases. A high precision means the proposed model did not miss true positives. Same is the case with recall scores. Precision and recall are combined to determine the F1 score. Therefore, this score is also high in all the classes. The model is able to classify well between correct and incorrect labelling of human actions. The accuracy of the proposed model is 95% which means that the model is performing well.

### 1) COHEN'S KAPPA COEFFICIENT

The agreement between two raters who each assign  $N$  items to  $C$  mutually exclusive categories is measured by Cohen's kappa [54]. The formula for  $K$  is

$$K = \frac{Pr(a) - pr(e)}{1 - Pr(e)} \tag{22}$$

where  $Pr(e)$  is the fictitious probability of chance agreement, calculated using the observed data to determine the odds of each observer arbitrarily choosing each category, and  $Pr(a)$  is the relative observed agreement among raters, often known as the total agreement probability.  $K = 1$  if there is total agreement among the raters.  $K = 0$  means that there is no agreement among the raters (except from what would be predicted by chance). The larger  $K$  means the classifiers are performing better. In this paper, Cohen's Kappa coefficient is used as classification performance measure. This metric essentially shows us how accurate the proposed model is than the classifier that simply makes random guesses based on the frequency of each class.

We have calculated Cohen's kappa score for the proposed model and the score come out to be  $K = 0.94$  which means that the model is performing well.

### 2) RECEIVER OPERATING CHARACTERISTIC CURVE (ROC CURVE)

ROC curve is an evaluation metric to measure multiclass performance. In general, there are two different methods for selecting multiple class performance indicators. OvR i.e., one versus rest and OVO i.e., one versus one. OvR stands for one way to assess multiclass models by contrasting every class simultaneously. In this case, we choose one class and think of it as our "positive" class, while the others are thought of as our "negative" class. OvO, which stands for "One vs One," is quite similar to OvR in that we compare all possible two-class combinations of the dataset rather than just one class against the other. In this paper, we have used

TABLE 6. Evaluation of performance metrics on the proposed model.

	Precision	Recall	F1-Score
Carry	0.95	0.97	0.96
Clap Hands	0.96	0.98	0.97
Pickup	1.00	0.95	0.98
Push	0.81	0.96	0.88
Sit down	0.94	0.97	0.95
Stand up	0.94	0.91	0.93
Throw	1.00	0.82	0.90
Walk	0.95	0.96	0.96
Wave Hands	0.98	0.99	0.98
<b>Micro Average</b>	<b>0.94</b>	<b>0.92</b>	<b>0.93</b>
<b>Weighted Average</b>	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>
<b>Accuracy</b>			<b>0.95</b>

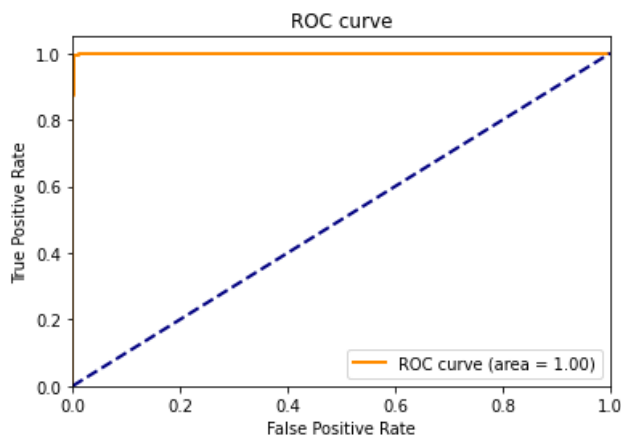


FIGURE 16. ROC curve for one class Vs rest of the classes).

OvR method for measuring the performance of the model. Figure 16 represents the ROC curve for one class versus rest of the classes. Area comes out to be 1 which means that the model is distinguishing between the classes efficiently.

The associated cumulative distribution functions should be  $F(p) = F(p(x)|0)$  and  $G(p) = G(p(x)|1)$ , respectively. A plot of  $G(p)$  on the vertical axis against  $F(p)$  on the horizontal axis is then used to define the ROC curve. This area undoubtedly falls within a unit square. An ROC curve in the upper left triangle of the square, which represents a good classification rule [55]. Figure 17 represents the ROC curve for all the classes.

## E. DISCUSSIONS

The suggested approach has created a customized multi-input feature fusion model for monitoring older people. The study employs handmade and self-learned techniques to recognize elderly daily routine behavior and forecast aberrant activity utilizing HOG and VGG-16 as primary and open datasets. Two separate features are taken into consideration as the prediction model is trained over an 80 percent dataset, and deep neural networks are used to recognize activity. The outcome analysis of the suggested model is discussed in this section. HOG + SVM is used in [37] to build human

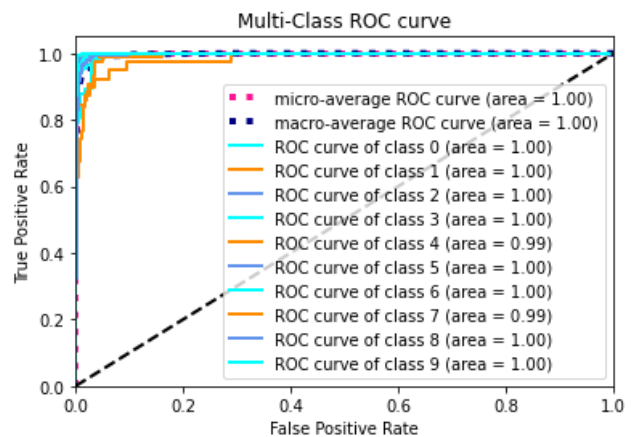


FIGURE 17. ROC curve for multi-class classification.

activity recognition system. The model was tested on open video dataset with 89.59% accuracy. We have also used HOG+SVM on our primary dataset and public dataset, the results were less accurate because of the noisy background in MSR Action Detection Dataset. Similarly, VGG-16 is used in [30] for indoor home scene recognition with an accuracy of 71%. When we used VGG-16 on our datasets, the accuracy scores are slightly better as shown in Table 7, but the reason behind less accuracy is the increase in number of feature dimensions resulting in dying ReLU. The proposed goldenAGER algorithm integrates handcrafted and self learnt features but feature fusion often leads to high dimensional features. goldenAGER uses a series of pooling layers and dense layers to reduce feature dimensionality which, in turn, save computational resources to train the model. Moreover, the existing models lack in finding strong intra-class correlation between activities. Our model uses spearman's correlation coefficient to find intra class variations between the activities of the same class. We have trained and tested our primary dataset and MSR action detection dataset on three different models. We have also compared our multi-input feature fusion model with VGG-16 model in Table 7. To design a HAR model using CNN, the 1000 features of VGG-16, as shown in Figure 7, are enhanced by using hidden layers

**TABLE 7. Accuracy of the three models on two different datasets.**

Model	Datasets	Accuracy
HOG + SVM	Primary Dataset	85.92%
	MSR Action Detection Dataset	71.17%
VGG-16	Primary Dataset	82.31%
	MSR Action Detection Dataset	75.84%
HOG+VGG-16	Primary Dataset	95%
	MSR Action Detection Dataset	93.08%

**TABLE 8. Performance comparison of proposed method with existing methods on MSR action detection dataset.**

Method	Recognition Accuracy
[56] SWS+ Rank based correlation method	90.20%
[57] Improved ELS	92.0%
[58] Baseline + MMDNN	87% , 92.88% respectively
[59] Multi-modal fusion	82%
<b>Proposed method</b>	<b>93.08%</b>

and dropout layers. These features are then integrated to a SoftMax layer consisting of 10 different classes. Three hidden layers connect the 1000 features in a manner similar to the second branch of the proposed model, as shown in Figure 12, providing 256 total features. To categorize 10 different actions, these 256 features are again fed to SoftMax. The evaluation of the proposed model in contrast to the VGG-16 CNN model explains the advancements made in fusing CNN and HOG characteristics. Figure 14 and Figure 15 show the accuracy graphs. In comparison to CNN model, proposed model extracts more detailed features. It enables the deep neural network to easily detect more complex patterns of different actions. As a result, the training of the proposed model becomes faster in each training cycle. A multi-input feature fusion model is created for action recognition in this paper. Using hand-engineered and self-learned feature descriptors, our proposed model can extract the features separately. At the same time, the filters in different channels are not shared, which ensures that each input can have a different convolutional setting for the different features. Each branch of the multi-input deep convolutional neural network proposed in this study can rely on the deep CNN to extract features, and the features extracted from each input will be concatenated together through the concatenate layer. This network can be thought of as a data fusion method at the feature level. The feature vector will then be forwarded to the network that is completely connected for feature fusion. Without relying on intricate feature engineering, this effective feature-level fusion technique may automatically fuse the most distinctive features from each channel and remove redundant features between channels. The entire procedure is end-to-end, which is better for subsequent real-time decision-making. Table 7 represents the accuracy score of the three models. It can be seen from the table that HOG+VGG-16 performs better as compared to using HOG and VGG-16 individually. The first model uses HOG as a feature descriptor and SVM for classification of the activity type. HOG+SVM gives accuracy 85.92% on primary dataset and 71.17% on MSR Action Detection Dataset. The accuracy is less because HOG can

extract the local features such as edges and corners from an image significantly but is unable to identify the complex patterns. Then, VGG-16 model is applied on both the datasets giving accuracy 82.31% and 75.84% respectively which is also comparatively less because it is not invariant to illumination changes like HOG. The combination of the two models gives better results for the primary dataset as well as publicly available dataset with accuracies 95% and 93.08% respectively. While HOG works on low level features, VGG-16 works on abstract level features to extract complex patterns from the image. Table 8 represents the performance comparison of our proposed method with the other existing methods on MSR action detection dataset. It can be seen clearly from the table that our method outperformed the other existing methods.

The proposed model can be used as a framework for many applications including continuous monitoring of elderly people. However, the model can only currently recognize a finite set of actions. As a result, the training data should be enhanced to include real-world activities for elderly.

## VIII. CONCLUSION

Automatic recognition and analysis of human activities especially elderly activities, is still a complicated task. The substantial intra-class correlation between various activities performed by senior people due to the age factor is the key problem with the design of such a system. We have used spearman's correlation coefficient to find the intra class correlation between the activities. The values for the 10 classes in the primary dataset i.e., Carry, Clap Hands, Pickup, Pull, Push, Sit Down, Stand Up, Throw, Walk and Wave Hands are +0.41, +0.59, -0.42, -0.08, -0.21, +0.77, +0.89, 0.91, -0.82 and +0.73 respectively. Activities such as Pick Up, Pull, Push, Throw and walk has strong intra class variation. HAR model's accuracy is affected by a number of problems, including illumination variations and variations in performing the same activity. The most recent CNN-based models perform admirably and increase accuracy; however, they fall short in differentiating between same actions. In order to



solve this issue, a personalized multi-input feature fusion algorithm, golderAGER that takes into account both hand-crafted and self-learned features is proposed in this study. To distinguish different human action patterns, VGG-16 and HOG features are extracted. This algorithm is trained and tested on real-time dataset which is collected from elderly volunteers with 10 different action classes. Our method performs well on this dataset with an accuracy of 95%. There exist multi-feature fusion models for the same application in the literature, but our proposed model is different in context of the model architecture, the optimization strategy used, loss function design, and the hyperparameters selection. The proposed method is compared with the existing methods summarizing the fact that our model aids in the development of strong HAR system to classify and recognize elderly activities in a more accurate way.

The proposed method recognizes only the simple activities that are mentioned above. This could be the major limitation of the proposed work. In the future, we intend to recognize more complex activities which contains human-human interaction or human-object interaction such as Handshake, hug, brushing teeth, cleaning the floor, cooking, watching TV etc. to check the validity of this model on these types of activities. We are also planning to add activities like falling to prevent life-threatening risks. We'll also compare the proposed model by fusing HOG features with other off-the-shelf CNN classifiers such as VGG-19, Wide-ResNet and ViT etc.

## ACKNOWLEDGMENT

The work of Diana Nagpal belongs to her Ph.D. study.

## CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

## REFERENCES

- [1] N. M. Nayak, R. J. Sethi, B. Song, and A. K. Roy-Chowdhury, "Modeling and recognition of complex human activities," in *Visual Analysis of Humans*. Berlin, Germany: Springer, 2011, pp. 289–309.
- [2] A. Hayat, F. Morgado-Dias, B. Bhuyan, and R. Tomar, "Human activity recognition in elderly people using machine and deep learning approaches," *Information*, vol. 13, no. 6, p. 275, May 2022.
- [3] S. Gupta and A. K. Saini, "Information system security and risk management: Issues and impact on organizations," *Global J. Enterprise Inf. Syst.*, vol. 5, no. 1, pp. 31–35, 2013.
- [4] T.-H.-C. Nguyen, J.-C. Nebel, and F. Florez-Revuelta, "Recognition of activities of daily living with egocentric vision: A review," *Sensors*, vol. 16, no. 1, p. 72, Jan. 2016.
- [5] N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, "Human activity recognition in artificial intelligence framework: A narrative review," *Artif. Intell. Rev.*, vol. 55, no. 6, pp. 4755–4808, 2022.
- [6] M. A. Gul, M. H. Yousaf, S. Nawaz, Z. Ur Rehman, and H. Kim, "Patient monitoring by abnormal human activity recognition based on CNN architecture," *Electronics*, vol. 9, no. 12, p. 1993, Nov. 2020.
- [7] D. Kumar, C. Verma, S. Dahiya, P. K. Singh, M. S. Raboaca, Z. Illés, and B. Bakariya, "Cardiac diagnostic feature and demographic identification (CDF-DI): An IoT enabled healthcare framework using machine learning," *Sensors*, vol. 21, no. 19, p. 6584, Oct. 2021.
- [8] S. Khater, M. Hadhoud, and M. B. Fayek, "A novel human activity recognition architecture: Using residual inception ConvLSTM layer," *J. Eng. Appl. Sci.*, vol. 69, no. 1, pp. 1–16, Dec. 2022.
- [9] D. Nagpal and R. Kumar, "Performance analysis of hybrid deep learning approach for human actions recognition," in *Proc. ICICCT*. Cham, Switzerland: Springer, 2022, pp. 173–188.
- [10] Y. Li and L. Wang, "Human activity recognition based on residual network and BiLSTM," *Sensors*, vol. 22, no. 2, p. 635, Jan. 2022.
- [11] P. U. Putra, K. Shima, and K. Shimatani, "A deep neural network model for multi-view human activity recognition," *PLoS ONE*, vol. 17, no. 1, Jan. 2022, Art. no. e0262181.
- [12] H. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, "Multi-view CNN-LSTM architecture for radar-based human activity recognition," *IEEE Access*, vol. 10, pp. 24509–24519, 2022.
- [13] 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.
- [14] S. Gupta and A. K. Saini, "An artificial intelligence based approach for managing risk of IT systems in adopting cloud," *Int. J. Inf. Technol.*, vol. 13, no. 6, pp. 2515–2523, Dec. 2021.
- [15] S. Gupta and A. Saini, "Cloud adoption: Linking business needs with system measures," *Global J. Enterprise Inf. Syst.*, vol. 9, no. 2, pp. 42–49, 2017.
- [16] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, "A review of human activity recognition methods," *Frontiers Robot. AI*, vol. 2, p. 28, Nov. 2015.
- [17] V. Nyemeesha and B. M. Ismail, "Implementation of noise and hair removals from dermoscopy images using hybrid Gaussian filter," *Netw. Model. Anal. Health Informat. Bioinf.*, vol. 10, no. 1, pp. 1–10, Dec. 2021.
- [18] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Proc. Comput. Sci.*, vol. 155, pp. 698–703, Jan. 2019.
- [19] A. Subasi, K. Khateeb, T. Brahimi, and A. Sarirete, "Human activity recognition using machine learning methods in a smart healthcare environment," in *Innovation in Health Informatics*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 123–144.
- [20] A. Kececi, A. Yildirak, K. Ozyazici, G. Ayluctarhan, O. Agbulut, and I. Zincir, "Implementation of machine learning algorithms for gait recognition," *Eng. Sci. Technol., Int. J.*, vol. 23, no. 4, pp. 931–937, Aug. 2020.
- [21] S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, "A review on human activity recognition using vision-based method," *J. Healthcare Eng.*, vol. 2017, pp. 1–31, Jan. 2017.
- [22] X. Wei, P. Wang, L. Liu, C. Shen, and J. Wu, "Piecewise classifier mappings: Learning fine-grained learners for novel categories with few examples," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6116–6125, Dec. 2019.
- [23] K. Muhammad, A. Ullah, A. S. Imran, M. Sajjad, M. S. Kiran, G. Sannino, and V. H. C. de Albuquerque, "Human action recognition using attention based LSTM network with dilated CNN features," *Future Gener. Comput. Syst.*, vol. 125, pp. 820–830, Dec. 2021.
- [24] C. Liu, J. Ying, H. Yang, X. Hu, and J. Liu, "Improved human action recognition approach based on two-stream convolutional neural network model," *Vis. Comput.*, vol. 37, no. 6, pp. 1327–1341, Jun. 2021.
- [25] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," 2016, *arXiv:1604.08880*.
- [26] J.-Y. He, X. Wu, Z.-Q. Cheng, Z. Yuan, and Y.-G. Jiang, "DB-LSTM: Densely-connected bi-directional LSTM for human action recognition," *Neurocomputing*, vol. 444, pp. 319–331, Jul. 2021.
- [27] J. Guo, M. Shi, X. Zhu, W. Huang, Y. He, W. Zhang, and Z. Tang, "Improving human action recognition by jointly exploiting video and WiFi clues," *Neurocomputing*, vol. 458, pp. 14–23, Oct. 2021.
- [28] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. García-Chamizo, "Gait analysis using computer vision based on cloud platform and mobile device," *Mobile Inf. Syst.*, vol. 2018, pp. 1–10, Jan. 2018.
- [29] Y. Cui, "Using deep machine learning to conduct object-based identification and motion detection on safeguards video surveillance," Brookhaven National Lab. (BNL), Upton, NY, USA, Tech. Rep. BNL-207942-2018-COPA, 2018.
- [30] A. Basu, L. Petropoulakis, G. D. Caterina, and J. Soraghan, "Indoor home scene recognition using capsule neural networks," *Proc. Comput. Sci.*, vol. 167, pp. 440–448, Jan. 2020.
- [31] M. Gholamrezai and S. M. T. Almodarresi, "Human activity recognition using 2D convolutional neural networks," in *Proc. 27th Iranian Conf. Electr. Eng. (ICEE)*, Apr. 2019, pp. 1682–1686.
- [32] I. Jegham, A. Ben Khalifa, I. Alouani, and M. A. Mahjoub, "Vision-based human action recognition: An overview and real world challenges," *Forensic Sci. Int., Digit. Invest.*, vol. 32, Mar. 2020, Art. no. 200901.

- [33] S. Khan, M. A. Khan, M. Alhaisoni, U. Tariq, H.-S. Yong, A. Armghan, and F. Alenezi, "Human action recognition: A paradigm of best deep learning features selection and serial based extended fusion," *Sensors*, vol. 21, no. 23, p. 7941, Nov. 2021.
- [34] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, Nov. 2019.
- [35] M. Kaur and S. Gupta, "Blockchain technology for convergence: An overview, applications, and challenges," in *Blockchain and AI Technology in the Industrial Internet of Things*. Hershey, PA, USA: IGI Global, 2021, pp. 1–17.
- [36] M. Kaur, S. Gupta, D. Kumar, C. Verma, B.-C. Neagu, and M. S. Raboaca, "Delegated proof of accessibility (DPoAC): A novel consensus protocol for blockchain systems," *Mathematics*, vol. 10, no. 13, p. 2336, Jul. 2022.
- [37] K. Seemanthini and S. S. Manjunath, "Human detection and tracking using HOG for action recognition," *Proc. Comput. Sci.*, vol. 132, pp. 1317–1326, Jan. 2018.
- [38] N. Zerrouki, F. Harrou, Y. Sun, and A. Houacine, "Vision-based human action classification using adaptive boosting algorithm," *IEEE Sensors J.*, vol. 18, no. 12, pp. 5115–5121, Jun. 2018.
- [39] Y. Hbali, S. Hbali, L. Ballihi, and M. Sadgal, "Skeleton-based human activity recognition for elderly monitoring systems," *IET Comput. Vis.*, vol. 12, no. 1, pp. 16–26, Feb. 2018.
- [40] R. Huan, C. Jiang, L. Ge, J. Shu, Z. Zhan, P. Chen, K. Chi, and R. Liang, "Human complex activity recognition with sensor data using multiple features," *IEEE Sensors J.*, vol. 22, no. 1, pp. 757–775, Jan. 2022.
- [41] S. Kiran, M. A. Khan, M. Y. Javed, M. Alhaisoni, U. Tariq, Y. Nam, R. Damasevicius, and M. Sharif, "Multi-layered deep learning features fusion for human action recognition," *Comput. Mater. Continua*, vol. 69, no. 3, pp. 4061–4075, 2021.
- [42] A. S. Chan, K. Saleem, Z. Bhutto, L. Memon, M. Hussain, S. A. Shaikh, and A. R. Siyal, "Feature fusion based human action recognition in still images," *Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 11, pp. 151–155, 2019.
- [43] L. Yao, Y. Liu, and S. Huang, "Spatio-temporal information for human action recognition," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, pp. 1–9, Dec. 2016.
- [44] T. Huynh-The, C. Hua, N. A. Tu, and D. Kim, "Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1533–1543, Feb. 2021.
- [45] A. Abdelbaky and S. Aly, "Two-stream spatiotemporal feature fusion for human action recognition," *Vis. Comput.*, vol. 37, no. 7, pp. 1821–1835, Jul. 2021.
- [46] P. Khaire, P. Kumar, and J. Imran, "Combining CNN streams of RGB-D and skeletal data for human activity recognition," *Pattern Recognit. Lett.*, vol. 115, pp. 107–116, Nov. 2018.
- [47] C. Zhang, Y. Xu, Z. Xu, J. Huang, and J. Lu, "Hybrid handcrafted and learned feature framework for human action recognition," *Appl. Intell.*, vol. 52, no. 11, pp. 12771–12787, 2022.
- [48] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 9–14.
- [49] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.
- [50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [51] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Questiones Geographicae*, vol. 30, no. 2, pp. 87–93, Jun. 2011.
- [52] Z. Ullah, L. Qi, D. Binu, B. R. Rajakumar, and B. M. Ismail, "2-D canonical correlation analysis based image super-resolution scheme for facial emotion recognition," *Multimedia Tools Appl.*, vol. 81, no. 10, pp. 13911–13934, Apr. 2022.
- [53] K. Akila and S. Chitrakala, "Highly refined human action recognition model to handle intraclass variability & interclass similarity," *Multimedia Tools Appl.*, vol. 78, no. 15, pp. 20877–20894, Aug. 2019.
- [54] S. M. Vieira, U. Kaymak, and J. M. C. Sousa, "Cohen's Kappa coefficient as a performance measure for feature selection," in *Proc. Int. Conf. Fuzzy Syst.*, Jul. 2010, pp. 1–8.
- [55] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, 2001.
- [56] M. Sharif, M. A. Khan, F. Zahid, J. H. Shah, and T. Akram, "Human action recognition: A framework of statistical weighted segmentation and rank correlation-based selection," *Pattern Anal. Appl.*, vol. 23, no. 1, pp. 281–294, Feb. 2020.
- [57] S. Wang, Z. Yu, and X. Yu, "Real-time online action detection and segmentation using improved efficient linear search," *Int. J. Comput. Sci. Math.*, vol. 10, no. 2, pp. 129–139, 2019.
- [58] Z. Bi and W. Huang, "Human action identification by a quality-guided fusion of multi-model feature," *Future Gener. Comput. Syst.*, vol. 116, pp. 13–21, Mar. 2021.
- [59] J. Liu, S. Song, C. Liu, Y. Li, and Y. Hu, "A benchmark dataset and comparison study for multi-modal human action analytics," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 2, pp. 1–24, May 2020.



**DIANA NAGPAL** received the B.Tech. degree in computer science and engineering from Punjab Technical University, Punjab, India, in 2009, and the M.Tech. degree from Guru Nanak Dev University, Amritsar, in 2011. She is currently pursuing the Ph.D. degree in machine learning with Chandigarh University, Gharuan. She is an Assistant Professor with the Guru Nanak Dev Engineering College, Ludhiana. She has been engaged in teaching and research in the field of digital image

processing. She has published many international journal and conference papers. Her research interests include artificial intelligence and machine learning.



**SHIKHA GUPTA** received the bachelor's and master's degrees in computer applications from Banasthali Vidyapith, Jaipur, Rajasthan, and the Ph.D. degree in computer applications from Teerthanker Mahaveer University, Moradabad, Uttar Pradesh. She is currently a Professor with the Department of Computer Science and Engineering, University Institute of Engineering, Chandigarh University, Gharuan, Punjab. She is also an Academician and a Researcher with extensive academic experience in prestigious institutes. She has been teaching for the past 25 years and has supervised several M.Tech./Ph.D. students. She has several research papers published in SCI/Scopus journals and conferences. She has also participated in and organized numerous FDPs, short-term courses, workshops, and seminars. Her research interests include artificial intelligence, blockchain technology, and machine learning.

She has also participated in and organized numerous FDPs, short-term courses, workshops, and seminars. Her research interests include artificial intelligence, blockchain technology, and machine learning.



**DEEPAK KUMAR** is currently a Professor with Chandigarh University, Mohali. He had also been the Ex-Director/Principal with the Bharat Group of Colleges, Sardulgarh, Mansa, Punjab. He has 16 years of teaching experience. He also successfully organized the Winning Camp (Technical Placement Preparation Event), from 4 July 2022 to 23 July 2022, with the Apex Institute of Technology, Chandigarh University. He has made significant contribution through various

approaches, such as directly supervising Ph.D. student, having acquired one patent, publishing more than 18 publications of research papers in international and national journals/conferences viz. MDPI and Springer. His research interest includes machine learning. He has been invited as a keynote speaker at various prestigious FDP in Akal University, Bathinda. He took parts in various international conference as the session chair.

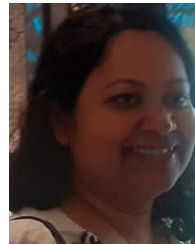


**ZOLTÁN ILLÉS** received the Ph.D. and Habilitation degrees in mathematics and physics from Eötvös Loránd University. Later, he took up the computer science supplementary course, which was started at that time. His Ph.D. dissertation was titled, “Implementation of Real-Time Measurements for High-Energy Ion Radiations,” in 2001. After graduation in 1985, he joined the Department of Computer Science, Eötvös Loránd University. In 2004, at the request of Jedlik Publisher, he also wrote a textbook on the C programming language. This book has a second, expanded edition, in 2008. In 2007, he was awarded a scholarship by the Slovak Academy of Sciences, where he spent six months in researching and teaching with Constantine the Philosopher University, Nitra. The NJSZT awarded the Rezső Tarján Prize, in 2016, for the success of the joint work that has been going on ever since. He and his colleagues also researched the issue of mobile devices and applications in the framework of a tender won, in 2014. Based on their research findings, he launched a pilot project to support real-time, innovative performance management. The first results of this research are an integral part of his habilitation dissertation. He got a Hungarian’s Republic Scholarship based on his outstanding academic achievements during his university studies. Since 2020, he has been an invited speaker at several international conferences and an Amity University Advisory Board Member.



**CHAMAN VERMA** received the Ph.D. degree in informatics from the Doctoral School of Informatics, Eötvös Loránd University, Budapest, Hungary, with Stipendium Hungaricum Scholarship funded by the Tempus Public Foundation, Government of Hungary. During his Ph.D. degree, he won the EFOP Scholarship, Co-Founded by the European Union Social Fund and the Government of Hungary, as a Professional Research Assistant in a real-time system, from 2018 to 2021. He also received the Stipendium Hungaricum Dissertation Scholarship of Tempus Public Foundation, Government of Hungary, from 2021 to 2022. He is currently an Assistant Professor with the Department of Media and Educational Informatics, Faculty of Informatics, Eötvös Loránd University. He has been awarded several Erasmus Scholarships for conducting international research and academic collaboration with European and non-European universities.

He received the best scientific publication awards from the Faculty of Informatics, Eötvös Loránd University, in the years 2021, 2022, and 2023. He has also been awarded ÚNKP Scholarship for research by the Ministry of Innovation and Technology and the National Research, Development and Innovation (NRDI) Fund, Government of Hungary, in 2020–2021 and 2022–2023. He has around ten years of experience in teaching and industry. He has more than 100 scientific publications in the IEEE, Elsevier, Springer, IOP Science, MDPI, and Walter de Gruyter. His research interests include data analytics, the IoT, feature engineering, real-time systems, and educational informatics. He is a Life Member of ISTE, New Delhi, India; a member of the editorial board; and a reviewer of various international journals and scientific conferences. He was the leading Guest Editor of the Special Issue *Advancement in Machine Learning and Applications* in mathematics, IF-2.25, MDPI, Basel, Switzerland, in 2022. He is a co-editor in the series of conference proceedings of ICRIC-2021 and ICRIC-2022 published by Springer, Singapore. He reviews many scientific journals, including IEEE, Springer, Elsevier, Wiley, and MDPI. He has Scopus citations of 615 with an H-index of 15. He has Web of Science citations 81 with an H-index of 8.



**BARNALI DEY** (Member, IEEE) received the B.Tech. and M.Tech. degrees in electronics and communication engineering and the Ph.D. degree in computational intelligence for cognitive radio from the National Institute of Technology Silchar, Assam, India, in 2017. Since 2021, she has been a Visiting Professor with the Aurel Vlaicu University of Arad, Romania. She is currently an Assistant Professor (Selection Grade) with the Sikkim Manipal Institute of Technology, Sikkim, India.

She has published several research articles in international journals and conferences of repute. Her research interests include machine learning (ML), artificial intelligence (AI) in cognitive radio, digital image processing, and advanced digital communications. She is a member of The Institution of Engineers, India. She is a regular reviewer of reputed journals indexed in SCOPUS/SCIE, including IEEE ACCESS in the field of AI, ML, and digital image processing.

• • •