



# Suspicious activity recognition for monitoring cheating in exams

Musa Dima Genemo<sup>1</sup>

Received: 27 July 2021 / Accepted: 7 February 2022 / Published online: 24 February 2022  
© Indian National Science Academy 2022

## Abstract

Video processing is getting special attention from research and industries. The existence of smart surveillance cameras with high processing power has opened the for making it conceivable to design intelligent visual surveillance systems. Know a day it is very possible to assure invigilators safety during the examination period. This work aims to distinguish the suspicious activities of students during the exam for surveillance examination halls. For this, a 63 layers deep CNN model is suggested and named "L4-BranchedActionNet". The suggested CNN structure is centered on the alteration of VGG-16 with added four blancheds. The developed framework is initially turned into a pre-trained framework by using the SoftMax function to train it on the CUI-EXAM dataset. The dataset for detecting suspicious activity is subsequently sent to this pre-trained algorithm for feature extraction. Feature subset optimization is applied to the deep features that have been obtained. These extracted features are first entropy coded, and then an ant colony system (ACS) is used to optimize the entropy-based coded features. The configured features are then input into a variety of classification models based on SVM and KNN. With a performance of 0.9299 in terms of accuracy, the cubic SVM gets the greatest efficiency scores. The suggested model was further tested on the CIFAR-100 dataset, and it was shown to be accurate to the tune of 0.89796. The result indicates the suggested frameworks soundness.

**Keywords** Suspicious activity · Cheating · Deep learning · Surveillance camera · Feature fusion

## Introduction

Examinations are a fundamental part of any education program. Academic cheating is usually administratively handled at the classroom or institutional level (Miller et al. 2017). We, as humans, possess an amazing skill to comprehend information that others pass on through their movements like the gesture of a certain body part or the motion of the entire body (Feng et al. Feb. 2017). Firstly, HAR is widely used in the area of smart homes (Du et al. 2019), where it plays two significant roles: health care of the elderly and disabled and adaptation of the environment to the residents' habits to improve the quality of their lives (Ternes et al. 2019). Now a day's HAR is widely used in every sector where security is the primary concern for the organization. For example, HAR is an application in the airport to detect passengers' activity around boarding areas, baggage claim, passenger

arrival places, city buses, university gates, and even in some cafeterias to recognize fighting, item lost, theft, vandalism, and the likes. To handle student suspicious activity in exam student cheating detection systems is in significant demand in universities during this COVID\_19 pandemic to ensure both safeties of the invigilators and students. Suspicious acts such as Back watching, Front watching, showing gestures, Side watching, Answer sharing, copying from notes, carrying a smartphone Walking in the exam hall warrant immediate attention. These activities demand a smart framework that can capable of issuing a warning or alarm as a result. A good Activity prediction can help the invigilator to take action for misbehaving of students in the exam. In addition, it limits the biases of invigilators while taking corrective measurements. HAR has the potential to have a significant impact on a variety of human activities. A deep learning algorithm can detect such behaviors automatically. (a) Simultaneous activities where individuals can complete a few activities simultaneously, like using scientific calculator for engineering students, using drawing equipment, (b) some times prediction ambiguity occurs, for example, front watching to copy from students who are in front of him or her can

✉ Musa Dima Genemo  
musa.ju2002@gmail.com

<sup>1</sup> Department of Computer Engineering, Gumushane University, Gümüşhane, Turkey

be related to normal front watching, (c) intraclass similarity, where students wear uniform cloth, (d) lack of lightning, (e) camera focus which is difficult to handle. Automating the surveillance events, reduce the invigilators' workload during an examination. Aside from its significance, the activity detection may worsen due to several technical obstacles. The following are some of the most significant challenges: (a) occlusion, (b) illumination, (c) variations in objects size, (d) variations in appearances because of varying clothing, (e) the computational time is another major challenge, and (f) the students may help each other by sharing pen or white paper, which makes the issue of distinguishing very hard to decide. In this work, a deep feature extraction methodology is presented for suspicious activity recognition to tackle the above-mentioned challenges. For this purpose, 63 layers of CNN-centered deep architecture are intended for feature acquisition. The acquired deep features are optimized through a feature selection algorithm. The foremost contributions are mentioned as under:

- (i) A dataset of different suspicious activities is prepared from CUI-EXAM and CIFAR-100 (Sawant 2018; Topirceanu 2017; Keresztury and Cser 2013)
- (ii) A 63 layers CNN network, named as L4-BranchedActionNet, is proposed. The network is initially trained with a CUI-EXAM dataset and the features of a suspicious activity recognition dataset are extracted on this pre-trained network.
- (iii) Entropy-coded ACS is applied for feature subset selection.
- (iv) Various classifiers are utilized to monitor the top classifier's functioning.
- (v) The outcomes depict the acceptable accomplishment of the intended work.

The manuscript is written according to the following sequence. “Introduction” and “Literature review” sections are comprised of the introduction part and the literature insight respectively. “Material and methods” section encompasses the proposed approach. “Results and discussion” section depicts results and discussion with details of Performance evaluation experiments and time graph. “Conclusion” section shows the paper's conclusions.

## Literature review

This section discusses the different methodologies used to detect human activity and suspicious actions in different literature. Convolutional Neural Networks (CNN) are used in a wide range of activities and have impressive results in a wide range of applications. The recognition of handwritten digits was one of the first applications in which CNN architecture was successfully applied (Jordan 2001). The addition

of new layers and the usage of other computer vision techniques have enhanced CNN networks steadily since its debut (George and Prakash 2018). Convolutional Neural Networks are mostly used in the ImageNet Challenge with various combinations of sketch datasets (Divakaran et al. 2018). Few researchers have compared the detection abilities of a human subject to those of a trained network using visual data. According to the results of the comparison, a human being corresponds to a 73.1% accuracy rate on the dataset, whereas the results of a trained network reveal a 64 percent accuracy rate (Mabrouk and Zagrouba 2018). Similarly, when Convolutional Neural Networks were applied to the same dataset, they achieved a 74.9% accuracy, exceeding human accuracy (Mabrouk and Zagrouba 2018). To achieve a significantly higher accuracy rate, the deployed approaches generally make use of the strokes' order. Studies are underway to better understand the behavior of Deep Neural Networks in a variety of circumstances (Booranawong et al. 2018). These experiments show how little adjustments to a picture can drastically alter grouping results. The work also includes images that are completely unrecognized by the public. There has been a lot of development in the area of feature detectors and descriptors and many algorithms and techniques have been developed for object and scene classification. We generally enticement the similarity between the object detectors, texture filters, and filter banks. Work is abundant in the literature on object detection and scene classification (Feng et al. Feb. 2017). Researchers mostly use the current up-to-date descriptors of Felzenszwalb and context classifiers of Hoeim (Ternes et al. 2019). The idea of developing various object detectors for basic interpretation of images is similar to the work done in multi-media communities in which they use a large number of “semantic concepts” for image and video annotations and semantic indexing (Hsu et al. 2018). In the literature that relates to our work, each semantic concept is trained by using either the image or frames of videos. As a result, with so many jumbled things in the scene, the technique is difficult to use and understand. Previous methods concentrated on single-object detection and classification based on human-defined feature sets. These proposed methods (Feng et al. Feb. 2017) investigate the relationship between objects in scene classification. The object bank was subjected to a variety of scene classification techniques to determine its utility. Many other forms of study have been done with an emphasis on low-level feature extraction for object detection and classification, such as the histogram of oriented gradient (HOG), GIST, filter bank, and a bag of feature (BoF) implemented through word vocabulary (Ternes et al. 2019). The various methodologies used in human activities detection and action recognition during the examination for detecting and classifying student activities during the exam. The activities are detected, recognized, and classified using a variety of

methods. Exam cheating is becoming a common occurrence around the globe, regardless of educational levels. To substantiate the conclusion, available research was examined. Various human-based activity approaches that provide a clue to conduct exam activity recognition research were defined in this section. According to the authors of this study, there is a very rare work in the literature that focuses on exam activity tracking using computer vision. However, this area can be associated with other human-based activities recognition frameworks. Despite the lack of specifically applicable work that can be used as a benchmark for this work, a thorough examination of various invigilation types and exam rules in higher education has been undertaken to solidify the study result. The student code of conduct books of ten (10) Ethiopian higher education institutions were examined to determine what acts are considered to be unlawful examination attempts. The invigilator assignment guidelines of selected universities (the two Ethiopian universities of science and technology, five from applied universities, and three from research universities) have been carefully investigated. The number of invigilators(s) is assigned depending on the number of students sitting for an exam. The whole(room) of the exam is also considered as input to allocate several invigilators. The review below illustrates the latest approaches found in the human activity categorization. Various researchers recognize human activity by various models like HMM model for shot boundary detection (He et al. 2018) as shown in Fig. 2.1 below adopted from HMM model for shot boundary detection Early studies on action recognition relied on hand-designed features and models (Chang et al. 2019; Hayes 2018). Recently various networks have been proposed to capture both the spatial and temporal information for video classification tasks including 2D CNN-based methods (Booranawong et al. 2018; Ranjan et al. 2019; Danielsson and Hansson 2018), RNN-based methods (Ternes et al. 2019), and 3D CNN-based methods (Du et al. 2019; Devine and Chin 2018; Noah et al. 2018; Venetianer et al. 2018; Ketcham 2017; Failed 2018c). In 2D CNN-based methods, high-level information is usually captured by a 2D CNN for each frame, and various fusion techniques including early and late fusion are applied to obtain the final prediction for each video. Various approaches are used in the related subfields such as for human detection(Nigam et al. 2019; Schwarzfischer et al. 2019; Finne et al. 2018), human behavior recognitions(Dhiman and Vishwakarma 2019; Hassan et al. 2018; Nigam et al. 2019; Schwarzfischer et al. 2019), and human activity detection(Jalal et al. 2019; Nguyen et al. 2018; Dhiman and Vishwakarma 2019). Zhang et al. (Sun et al. 2018) propose a new technique detecting the human in RGB-D pictures that integrate region of Interest(ROI) creation, depth size relationship approximation, and human indicator. Several authors also address different neural network models. In 2D CNN-based methods (Tripathi et al. 2018;

Irfan et al. 2018; Zhang 2014), high-level information is captured by a 2D CNN for each frame, and various fusion techniques including early and late fusion are applied. Some interesting analytical studies have recently investigated which categorize of videos require temporal information for recognition. The approaches in Jalal et al. (2019); Nguyen et al. 2018) utilize the video classification 14 methods that exploit the appearance information of the object of interest in video to produce a highly accurate 3D classification. Given limited annotated from only 20%-50% annotated samples, the proposed approach can learn CNNs that can potentially outperform those trained in a fully supervised manner. Xin et al. (Dhiman and Vishwakarma 2019; Hassan et al. 2018) have performed improvement in the global features. The improvement over the supervised method, details for the generation of natural language explanations in addition to visual information in the video allows for further analysis of video labeling. Also, there exists a lot of work in human recognition with a vital role in activity recognition(Ranjan et al. 2019; Agarwal et al. 2019).

## Material and methods

The materials and implementation techniques used in this work will be discussed in this section in detail. Also, the explanation of the suggested 63 layers CNN model is discussed in this section (Fig. 1). The framework's major steps include Data preparation, Handcrafted data labeling, Training the proposed CNN architecture on the CUI-EXAM dataset, feature extraction of the action recognition dataset on the proposed CNN architecture, feature subset selection using (ACS) algorithm, and classification using various classifiers. In an autonomous features extraction and classification pipeline, a new CNN-based model with 63 layers is proposed. L4-Branched-ActionNet is the name given to the entire pipeline. Figure 2 depicts the proposed L4-BranchedActionNet's graphical structure. VGG16 [63] serves as the foundation for the proposed architecture pipeline. Finally, the ensemble selected feature vector is fed to the SVM-based classifiers to get the classification results. The detailed block diagram is shown in Fig. 2. The steps depicted in the block diagram are discussed one after the other in the accompanying section. It is very important to label students' suspicious activities during the examination. Although many researchers pay high attention to HAR surveillance security to ensure human safety and health as well.

L4-Branched-ActionNet was introduced mainly to enable the training process across clustered GPUs with low memory capacity. The filters are split into multiple divisions in a Conv. All groups oversee a collection of 2D convolutions with a specific range. And Batch Norm is used for adjusting channel neurons over a small batch's defined amount. It



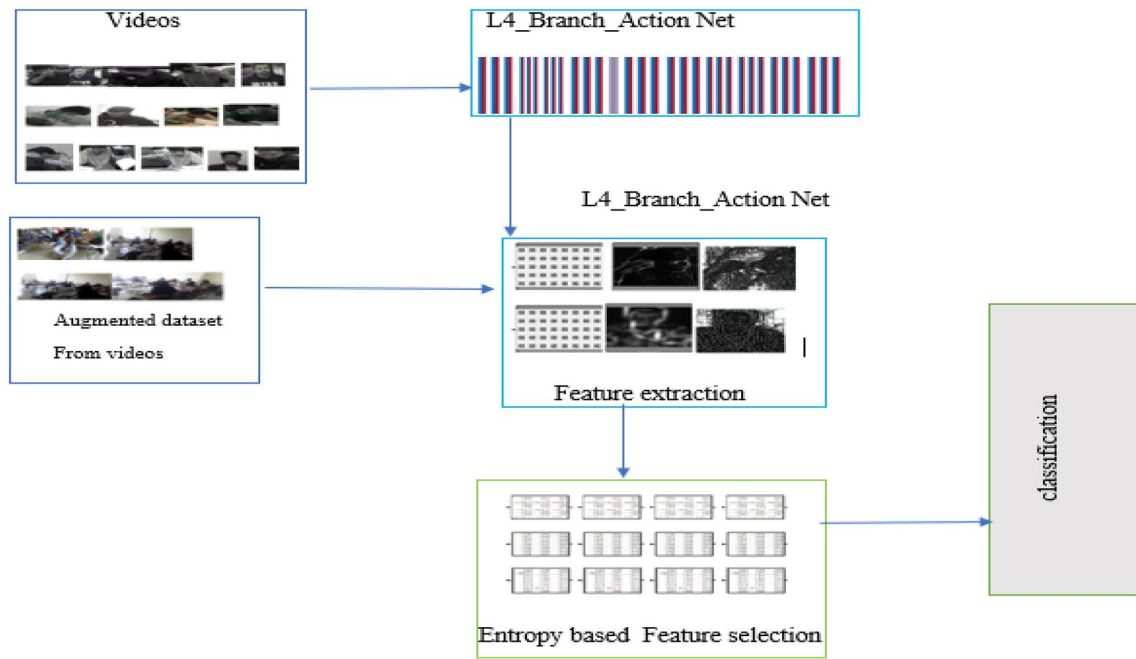


Fig. 1 The intuition of the proposed L4-Branched-ActionNet based framework for suspicious activity recognition



Fig. 2 Structure of proposed framework

calculates the mean and variance in fragments. The mean is derived, and the features are separated using the standard

deviation. The mean of the batch  $\mathbb{B} = \mathbb{I}_z \cdots \mathbb{I}_w$  is measured as follows:



$$MeanB = 1/w \sum_z^w 1||z \tag{1}$$

here  $w$  represents the number of feature maps in a batch. The proposed CNN model employs both ReLU and Leaky\_ReLU operations. The standard ReLU transforms all numbers that are lower than 0 to 0. For values less than zero, Leaky ReLU has a small slope rather than being zero.

### Methods

The framework proposes a classification of students' suspicious activity in the exam hall. Labeling of such activities is based on suspicious activities taken as input from surveillance cameras during the examination. The proposed methodology is developed for the system that is based on computer vision. The model integrates resizing the images in the whole dataset and their conversion to grayscale images, feature extraction, selection, and classification of images. Afterward, the fused features are selected by implementing the Principal Component Analysis method. Lastly, the selected features are classified using a support vector machine and fine KNN. The proposed method uses the newly created dataset to evaluate its effectiveness.

### Deep feature extraction

The proposed approach is intended to feature extraction from the deep-trained CNN pipeline. Therefore, for Training a CUI-EXAM is employed. There are 3500 Training and 500 validation images for each class. All the learning and validation images are mixed for pre-training, making 600 images in every class. The mixed dataset is supplied to the proposed CNN model for training. The trained network is then used for feature extraction on action recognition datasets and the FC\_18 layer is chosen for features extraction. Total 4096 features are attained per image from the FC\_18 layer. The prepared dataset contains a total of 4000 images extracted from videos for training. This makes the feature set dimension of all datasets  $4000 \times 4096$ .

### Feature selection

For feature selection, we use the PCA algorithm to select the robust features and to elicit the selected featured subset from the set of complete feature sets. In the cheating dataset, it is very tough to identify an individual parameter that further characterizes the performance of matching across the number of features(Noah et al. 2018).

### Ant colony optimization-based features subset selection

This approach comes under the category of wrapper-based approach. It is also known as ant colony system-based feature optimization. It depends on the probability theory (Finne et al. 2018). It is inspired by ants' behaviors. The ants while traveling, spread a substance known as a pheromone. This substance intensity reduces with time. The ants follow the route with probabilistically high intensity of pheromone. This guides them to seek the least cost route. The movement of ants is just like traveling in a graph i.e., from node to node. A node depicts a feature and the edges among the nodes show the option to select the following feature. The algorithm seeks the optimal features. The algorithm stops when the least number of nodes are visited, and a stopping condition is reached. All nodes in the graph are connected in a mesh-like structure. The pheromone values are coupled with nodes (features). The feature is selected by an ant depending on the probability at a certain time mathematically given as:

$$p_j^m(\uparrow) = \begin{cases} \frac{|\Gamma_j(\uparrow)|^{\bar{w}} |\rho_j|^W}{\sum_{v \in (\xi_1, \xi_2, \dots, \xi_m)} |\Gamma_v(\uparrow)|^{\bar{w}} |\rho_v|^W} & \text{if } j \in (\xi_1, \xi_2, \dots, \xi_m) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where  $(\xi_1, \xi_2, \dots, \xi_k)$  represents the feature set. If these features are not yet visited, then they will be the component of the unfinished solution.  $\Gamma_j(\uparrow)$  and  $\rho_j$  depict pheromone and empirical values linked with the  $i$ th feature.  $\bar{w}$  and  $W$  show the cost of pheromone and empirical knowledge, respectively.  $\uparrow$  portrays the time limit.

### Feature fusion

If we describe SFTA feature vectors as:

$$F_{S_{1 \times d}} = \{ S_{1 \times 1}, S_{1 \times 2}, S_{1 \times 3} \dots S_{1 \times d} \} \tag{3}$$

where  $F_{S_{1 \times d}}$  represents the vector dimensions of the SFTA feature vector, Furthermore LBP feature vector describes as:

$$F_{L_{1 \times c}} = \{ L_{1 \times 1}, L_{1 \times 2}, L_{1 \times 3} \dots L_{1 \times c} \} \tag{4}$$

where  $F_{L_{1 \times d}}$  represents the vector dimensions of the LBP feature vector, Furthermore Gabor feature vector describes as:

$$F_{G_{1 \times r}} = \{ G_{1 \times 1}, G_{1 \times 2}, G_{1 \times 3} \dots G_{1 \times r} \} \tag{5}$$

where  $F_{G_{1 \times d}}$  represents the vector dimensions of the Gabor feature vector, the given input features vectors are simply linked together successively or horizontally. The concatenated vector is represented as:

$$F_{R_{1 \times j}} = \sum (F_{S_{1 \times d}}, F_{L_{1 \times c}}, F_{G_{1 \times r}}) \tag{6}$$



where  $F_{R_{1 \times j}}$  the output resultant vector is the combined feature vector after concatenation and  $j = d + c + r$  then the mean is calculated of the output feature vector as:

$$F_k = \frac{1}{j} \sum_{i=0}^j F_{R_i} \quad (7)$$

Following that Euclidian distance is searched out for each feature w.r.t mean value and put on an activation function that will organize the vector in the least distance order. The Euclidian distance of the features are explained as:

$$F_d = \{d_1, d_2, d_3, \dots, d_j\} \quad (8)$$

The resultant feature vector is represented as:

$$f_{F_{1 \times k}} = \{F_{1 \times 1}, F_{1 \times 2}, F_{1 \times 3}, \dots, F_{1 \times K}\} \quad (9)$$

## Classification

Next after extracting the feature, the system is classified utilizing different classifiers and a verification process is implemented based on these selected features. In this phase of activities classification, we apply the selected classifiers Fine KNN and Cubic SVM. The entropy-coded ACS-based chosen features are at the end passed to the predictor for categorization. The various SVM versions (support vector machine) and KNN (K nearest neighbors) are deployed to observe the system performance. Such versions include linear SVM (Lin-SVM), quadratic SVM (Quad-SVM), fine Gaussian SVM (Fin-Gaus-SVM), medium Gaussian SVM (Med-Gaus-SVM), coarse Gaussian SVM (Cor-Gaus-SVM), cubic SVM (Cub-SVM), cosine KNN (Cos-KNN), coarse KNN (Cor-KNN), and fine KNN (Fin-KNN). The detailed study of SVM versions can be retrieved from Schwarzfischer et al. (2019); Noah et al. 2018; Li et al. 2018) while the detailed study of KNN versions can be regained from Booranawong et al. (2018); Mabrouk and Zagrouba 2018; Hsu et al. 2018). Observing the performance outcomes, Cub-SVM becomes the best-performed classifier for the selected action datasets. I evaluated different classification algorithms (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, and nearest-neighbors), implemented in MATLAB. I used 4000 datasets, which represent the whole CUI\_Exam dataset to achieve significant conclusions about the classifier behavior. The classifiers most likely to be the bests are the support vector machine (SVM) versions, the best of which (implemented in MATLAB) achieves 0.9299 of the maximum accuracy.

## Results and discussion

In this work, the proposed architecture has been described. The main goal of this research is to create a CNN structure that can tackle the supplied dataset. The Deep L4-BranchedActionNet CNN Network suggested here is solely utilized to extract powerful features following feature selection. The pretraining is performed on a third-party dataset i.e., CUI\_EXAM. Also, the proposed 63 layers Deep L4-Branched-ActionNet CNN Network is created after extensive experimentation. After many experiments performed on six classes (Front watching, Back watching, Side watching, Normal, Suspicious watching, and Showing gestures). Different approaches are followed to finalize this architecture. The foremost approaches include fine-tuning, adding, and removing different layers. In its final form, the 63 layers architecture is found good with the best outcomes in terms of performance. The discussion and interpretation of the outcomes of the proposed framework are described in this portion. The dataset is defined first in this section. The procedure for evaluating performance is then portrayed. Finally, the tests are well discussed. All the mentioned experiments in this manuscript are conducted on a Pentium core i-5 system containing 8 GB of memory. The training is aided with NVIDIA GTX 1070 GPU comprising 8 GB RAM. The coding is performed by using MATLAB2020a.

### Dataset (preparation and augmentation)

The training and performance evaluation in this work is performed on two different datasets. The first dataset is the **CUI-EXAM** which is comprised of approximately 4000 images of students 'activities acquired through CCTV cameras during exams. The second dataset is CIFAR-100. CIFAR is a repository of images with 100 classes. Dataset (CUI-EXAM) is for students behavior detection and classification containing bounding boxes from 100 videos annotated into 6 action classes (Front watching, Back watching, Side watching, Normal, Suspicious watching, and Showing gestures) (Fig. 3). To increase the size of the dataset, image augmentation is utilized by flipping the images. This doubles the size of the dataset. The sample dataset is shown in Fig. 4.

### Performance measures

The 5-folds mechanism for the cross-validation is employed for booth learning and assessment. fivefold cross-validation is used for ground reality class marks. It makes up 80% of the data for each fold is chosen at random for preparation, while the remaining 20% is chosen at random for testing. To evaluate the quality of the algorithm, some evaluation

measurements such as Accuracy, Confusion matrix, Misclassification rate, TPR, FPR, Sensitivity (SEN), specificity (SPE), precision (PRE), prevalence, Error rate (ER), F Score and ROC Curves, and AUC are calculated from output image. First, the algorithm is implemented on the given input image taken from the dataset of cheating monitoring during the examination to extract the output image or feature image. The evaluation metric is then calculated to check the efficiency of a given algorithm. Sensitivity and specificity

were calculated to assess the correctness of the algorithm (Fig. 5). Higher accuracy scores, SEN, and SPE ensure the classifier’s effectiveness also the good quality of the algorithm. Other evaluation measures can ensure the error rate or erroneous results obtained by a given algorithm. TPR and TNR are used to obtain the true classification rate of a given algorithm. The value of TPR and TNR or 1 value ensures that the given algorithm produces an error-free and better result. FPR and FNR are used to obtain the error rate

Fig. 3 Representation of a Datasets

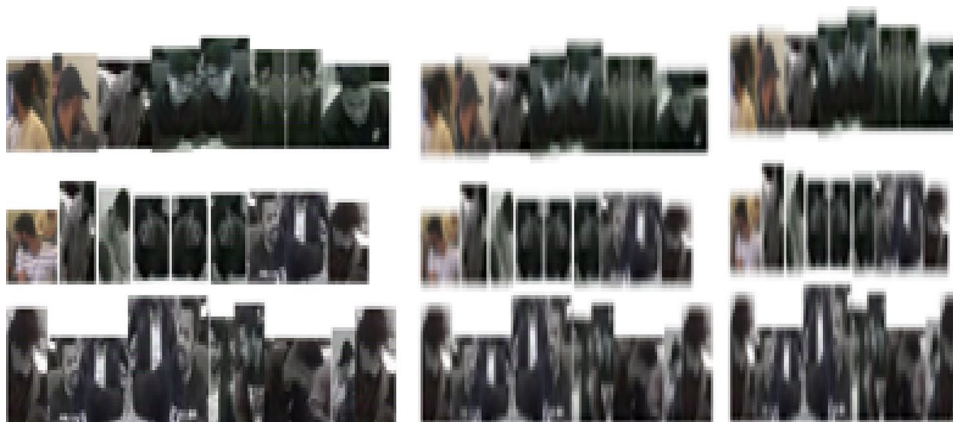


Fig. 4 CUI-EXAM Dataset

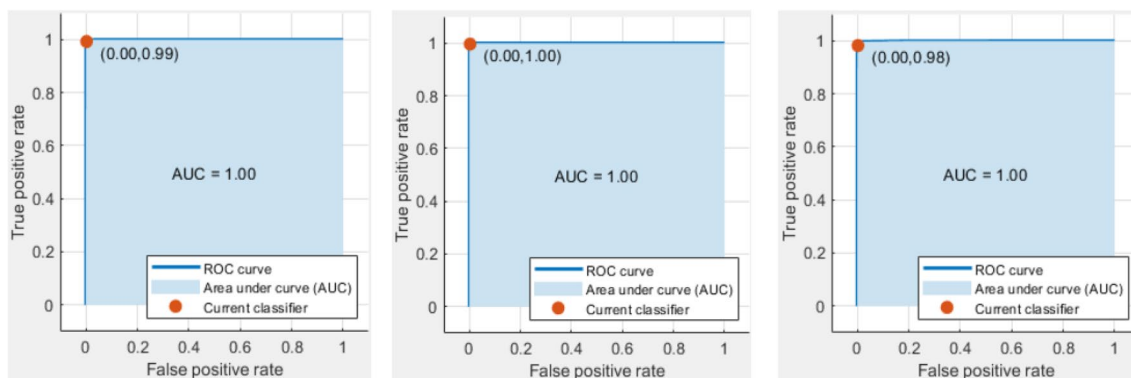
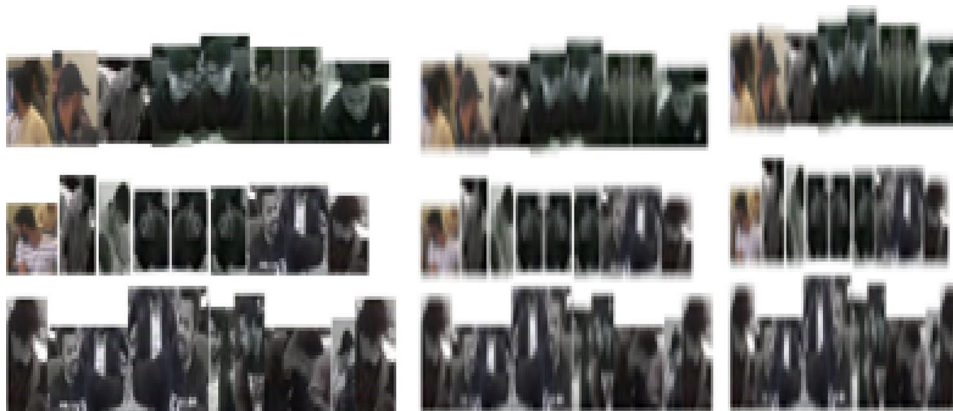


Fig. 5 ROCs and AUCs of all classes having best results using CSVM on CUI-Exam dataset



**Table 1** Performance of the model on the datasets

| Classifier | Correctness | Sensitivity | Specificity | Precision | Measure | Percent |
|------------|-------------|-------------|-------------|-----------|---------|---------|
| LSVM       | 64.53       | 73.28       | 62.62       | 29.94     | 42.52   | 67.74   |
| QSVM       | 83.22       | 91.01       | 81.53       | 51.79     | 66.01   | 86.14   |
| FGSVM      | 42.68       | 47.29       | 41.68       | 15.02     | 22.80   | 44.39   |
| MGSVM      | 83          | 90.52       | 82.25       | 52.56     | 66.58   | 86.28   |
| CGCVM      | 73.07       | 58.37       | 54.35       | 21.80     | 31.75   | 56.33   |
| CSVM       | 92.99       | 94.33       | 89.69       | 66.61     | 78.08   | 92.99   |

of a given algorithm. The value of FPR and FNR or zero value ensures the correctness of the given algorithm. Every point on the ROC curve shows the SEN/SPE correspondence towards a particular decision threshold. AUC measures the extent to which the parameters distinguish between different classes.

$$\text{Sensitivity (recall/TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (11)$$

$$\text{Specificity(TNR)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (12)$$

$$\text{AUC(TNR)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

$$\text{F-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

$$\text{G-Mean} = \sqrt{\text{TPR} \times \text{TNR}} \quad (16)$$

### Training/testing procedure

For Performance evaluation of the proposed technique, 4000 images of students' (during the examination) database are trained and tested, these images are trained on all SVM's (Support vector machine) and KNN's classifier learners including all SVM classifiers and KNN classifiers. Results of this performance evaluation against the proposed technique of feature fusion, there are several experiments performed using different features both in,

individual and combination of multiple features for the evaluation of best feature combination based on their high Accuracy score. This results in us the best combination of feature descriptors of the highest returning Accuracy score of the classifiers of Cubic SVM and Fine KNN.

### Conclusion

Suspicious activity recognition has been an important area of research in recent years. Recognizing students' suspicious activities automatically in a well-timed manner will help the invigilator to take accurate and fair action. This work is encompassed to classify the suspicious activities using a proposed 63 layers CNN network named L4-Branched-ActionNet. The network is trained on the CUI-EXAM dataset. The features are then fed to an entropy-coded ACS scheme to reduce the features. The training and testing of the dataset selected features are performed with different variation versions of SVM and KNN categorizers. The findings are repeated on these classifiers by altering the number of features at the feature choice phase. The lower performance is attained on 100 features with an accuracy of 0.9299 with the Cub-SVM classifier. The best classification results are considered with 1000 features using a Cub-SVM classifier having an accuracy of 0.9299. The CSVM is found to be the overall best, having better performance in all experiments. The results are also validated on the CIFAR-100 dataset and compared with recent works. The acceptable and comparable results demonstrate the legitimacy of the suggested approach. Feature fusion can be implemented by taking features from another CNN-based network. Existing works show superior outcomes in this regard. However, we suggest this task be explored in the upcoming future. Moreover, new deep learning building blocks and feature selection methods can be checked in this domain for a dominant performance as to future work (Tables 1, 2 and 3).



**Table 2** CUI-EXAM Dataset detail

| Dataset  | Back watching | Front watching | Normal | Showing gesture | Side watching | Suspicious |
|----------|---------------|----------------|--------|-----------------|---------------|------------|
| Original | 407           | 286            | 561    | 327             | 380           | 313        |
| Flipped  | 300           | 210            | 350    | 200             | 250           | 250        |

**Table 3** Confusion matrix of CSVM classifier on CUI-Exam dataset

| Backwatching   | 0.92          | 0.00           | 0.01   | 0.02             | 0.01          | 0.01       |
|----------------|---------------|----------------|--------|------------------|---------------|------------|
| Front watching | 0.00          | 0.91           | 0.00   | 0.01             | 0.00          | 0.01       |
| Normal         | 0.01          | 0.01           | 0.90   | 0.00             | 0.01          | 0.00       |
| Showinggesture | 0.00          | 0.00           | 0.01   | 0.91             | 0.00          | 0.01       |
| Side watching  | 0.00          | 0.00           | 0.02   | 0.01             | 0.90          | 0.00       |
| Suspicious     | 0.01          | 0.00           | 0.00   | 0.00             | 0.01          | 0.92       |
|                | Back watching | Front watching | Normal | Showing watching | Side watching | Suspicious |

**Declarations**

**Conflict of interest** Hence the Author is one person, there is no possibility of conflict of interest at all. No organization sponsored this work. Everything is covered by the researcher.

**References**

Abdulghani, H.M., Haque, S., Almusalam, Y.A., Alanezi, S.L., Alsulaiman, Y.A., Irshad, M., et al.: Self-reported cheating among medical students: an alarming finding in a cross-sectional study from Saudi Arabia. *PLoS ONE* **13**, e0194963 (2018)

Agarwal, R. Jain, R. Regunathan, R., Kumar, C. P.: Automatic attendance system using face recognition technique. In: *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, pp. 525–533 (2019)

Alizadeh, M., Peters, S., Etalle, S., Zannone, N.: Behavior analysis in the medical sector: theory and practice. In: *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pp. 1637–1646 (2018)

Aravena, C., Vo, M., Gao, T., Shiratori, T., Yu, L.-F., Contributors, E.: Perception meets examination: Studying deceptive behaviors in VR. In: *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (2017)

Aristizabal, D.A., Denman, S., Nguyen, K., Sridharan, S., Dionisio, S., Fookes, C.: Understanding patients' behavior: vision-based analysis of seizure disorders. In: *IEEE Journal of Biomedical and Health Informatics*, (2019a)

Baradel, F., Wolf, C., Mille, J., Taylor, G. W.: Glimpse clouds: Human activity recognition from unstructured feature points. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 469–478 (2018)

Ben-Musa, A.S., Singh, S.K., Agrawal, P.: Suspicious human activity recognition for video surveillance system. *ICCICCT* (2014)

Booranawong, A., Jindapetch, N., Saito, H.: A system for detection and tracking of human movements using RSSI signals. *IEEE Sens. J.* **18**, 2531–2544 (2018)

Brewer, P. C., Eubanks, D., Gupta, H., Scanlon, W. A., Venetianer, P. L., Yin, W. et al.: Object tracking and best shot detection system. Google Patents (2018)

Chang, P.D., Wong, T.T., Rasiej, M.J.: Deep learning for detection of complete anterior cruciate ligament tear. *J. Digit. Imaging* **44**, 1–7 (2019)

Danielsson, N., Hansson, A.: Method and system for tracking an object in a defined area. Google Patents (2018)

Desai, N., Pathari, K., Raut, J., Solavande, V.: Online surveillance for exam. *Jung*, vol. 4, (2018a)

Devine, C.A., Chin, E.D.: Integrity in nursing students: a concept analysis. *Nurse Educ. Today* **60**, 133–138 (2018)

Dhiman, C., Vishwakarma, D.K.: A review of state-of-the-art techniques for abnormal human activity recognition. *Eng. Appl. Artif. Intell.* **77**, 21–45 (2019)

Divakaran, A., Yu, Q., Tamrakar, A., Sawhney, H. S., Zhu, J., Javed, O. et al.: Real-time object detection, tracking and occlusion reasoning. Google Patents (2018)

Du, Y., Lim, Y., Tan, Y.: A novel human activity recognition and prediction in smart home based on interaction. *Sensors* **19**(20), 4474 (2019)

Elias, S.J., Hatim, S.M., Hassan, N.A., Latif, L.M.A., Ahmad, R.B., Darus, M.Y., et al.: Face recognition attendance system using local binary pattern (LBP). *Bull. Electr. Eng. Inform.* **8**, 12 (2019b)

Feng, S., Setoodeh, P., Haykin, S.: Smart home: cognitive interactive people-centric internet of things. *IEEE Commun. Mag.* **55**(2), 34–39 (2017)

Finne, E., Glausch, M., Exner, A.-K., Sauzet, O., Stoelzel, F., Seidel, N.: Behavior change techniques for increasing physical activity in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. *Cancer Manag. Res.* **10**, 5125 (2018)

George, R.P., Prakash, V.: Real-time human detection and tracking using quadcopter. In: *Intelligent Embedded Systems*, Springer, pp. 301–312 (2018)

Hassan, M.M., Huda, S., Uddin, M.Z., Almgren, A., Alrubaian, M.: Human activity recognition from body sensor data using deep learning. *J. Med. Syst.* **42**, 99 (2018)

Hayes, A.L.: Autism spectrum disorder: patient care strategies for medical imaging. *Radiol. Technol.* **90**, 31–47 (2018)

He, H., Zheng, Q., Li, R., Dong, B.: Using face recognition to detect “Ghost Writer” cheating in examination. In: *International Conference on E-Learning and Games*, pp. 389–397 (2018)



- Hsu, S.-C., Chuang, C.-H., Huang, C.-L., Teng, R., Lin, M.-J.: A video-based abnormal human behavior detection for psychiatric patient monitoring. *Int Worksh. Adv. Image Technol. (IWAIT)* **2018**, 1–4 (2018)
- Irfan, M., Tokarchuk, L., Marcenaro, L., Regazzoni, C.: Anomaly detection in crowds using multi sensory information. In: *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6 (2018)
- Jalal, A., Maria, M., Siddiqi, M.: Robust Spatio-temporal features for human interaction recognition via an artificial neural network. In: *IEEE Conference on International Conference on Frontiers of information technology*, (2018)
- Jalal, A., Mahmood, M., Hasan, A.S.: Multi-features descriptors for human activity tracking and recognition in Indoor-outdoor environments. In: *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 371–376 (2019)
- Jordan, A.E.: College student cheating: The role of motivation, perceived norms, attitudes, and knowledge of institutional policy. *Ethics Behav.* **11**, 233–247 (2001)
- Keresztury, B., Cser, L.: New cheating methods in the electronic teaching era. *Procedia Soc. Behav. Sci.* **93**, 1516–1520 (2013)
- Kerkvliet, J., Sigmund, C.L.: Can we control cheating in the classroom? *J. Econ. Educ.* **30**, 331–343 (1999)
- Ketcham, M.: CCTV Face Detection Criminals and Tracking System Using Data Analysis Algorithm. In: *Advances in Intelligent Informatics, Smart Technology, and Natural Language Processing: Selected Revised Papers from the Joint International Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP 2017)*, p. 105 (2019)
- Lewis, M.A., Neighbors, C.: An examination of college student activities and attentiveness during a web-delivered personalized normative feedback intervention. *Psychol. Addict. Behav.* **29**, 162 (2015)
- Li, K., He, F.-Z., Yu, H.-P.: Robust Visual Tracking based on convolutional features with illumination and occlusion handling. *J. Comput. Sci. Technol.* **33**, 223–236 (2018)
- Mabrouk, A.B., Zagrouba, E.: Abnormal behavior recognition for intelligent video surveillance systems: a review. *Expert Syst. Appl.* **91**, 480–491 (2018)
- Manoharan, S.: Cheat-resistant multiple-choice examinations using personalization. *Comput. Educ.* **130**, 139–151 (2019)
- Miller, A.D., Murdock, T.B., Grotewiel, M.M.: Addressing academic dishonesty among the highest achievers. *Theory into Pract.* **56**, 121–128 (2017)
- Van Nguyen, N.H., Pham, M.T., Ung, N.D., Tachibana, K.: Human activity recognition based on weighted sum method and combination of feature extraction methods. *Int. J. Intell. Inform. Syst.* **7**, 9 (2018)
- Nguyen, P.H., Turkey, C., Andrienko, G., Andrienko, N., Thonnard, O., Zouaoui, J.: Understanding user behavior through action sequences: from the usual to the unusual. *IEEE Trans. Visualization Comput. Graph.* **25**(9), 2838–2852 (2018c)
- Nigam, S., Singh, R., Misra, A.: Towards intelligent human behavior detection for video surveillance. In: *Censorship, Surveillance, and Privacy: Concepts, Methodologies, Tools, and Applications*, IGI Global, pp. 884–917 (2019)
- Noah, B., Keller, M.S., Mosadeghi, S., Stein, L., Johl, S., Delshad, S., et al.: Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. *NPJ Digital Med.* **1**, 2 (2018)
- Ranjan, R., Patel, V.M., Chellappa, R.: Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 121–135 (2019)
- Sawant, M.R.: Emerging trends in ethical behaviour in education industry. *Emerg Trends Manag New Perspect Pract* **24**, 357 (2018)
- Schwarzfischer, P., Gruszfeld, D., Stolarczyk, A., Ferre, N., Escribano, J., Rousseaux, D., et al.: Physical activity and sedentary behavior from 6 to 11 years. *Pediatrics* **143**, 20180994 (2019)
- Singh, A.K., Bansal, V.: SVM Based approach for multiface detection and recognition in static images. *J. Image Process. Artif. Intell.* **4**, 17 (2018b)
- Sun, X., Wu, P., Hoi, S.C.: Face detection using deep learning: an improved faster RCNN approach. *Neurocomputing* **299**, 42–50 (2018)
- Sung, J., Ponce, C., Selman, B., Saxena, A.: Unstructured human activity detection from RGB images. In: *2012 IEEE International Conference on Robotics and Automation*, pp. 842–849 (2012)
- Ternes, M., Babin, C., Woodworth, A., Stephens, S.: Academic misconduct: an examination of its association with the dark triad and antisocial behavior. *Personal. Individ. Differ.* **138**, 75–78 (2019)
- Topirceanu, A.: Breaking up friendships in exams: a case study for minimizing student cheating in higher education using social network analysis. *Comput. Educ.* **115**, 171–187 (2017)
- Tripathi, R.K., Jalal, A.S., Agrawal, S.C.: Suspicious human activity recognition: a review. *Artif. Intell. Rev.* **50**, 283–339 (2018)
- Venetianer, P. L., Lipton, A. J., Chosak, A. J., Frazier, M. F., Haering, N. Myers, G. W. et al.: Video surveillance system employing video primitives. Google Patents (2018)
- Zhang, H.: 3D robotic sensing of people: human perception, representation and activity recognition. (2014)
- Zhigang, D., Guangxue, D., Huan, L., Guangbing, Z., Nan, W., Wenjie, Y.: Human behavior recognition method based on double-branch deep convolution neural network. In: *2018 Chinese Control And Decision Conference (CCDC)*, pp. 5520–5524 (2018)