Design and Implementation of Human Safeguard Measure Using Separable Convolutional Neural Network Approach

R. Vaitheeshwari, V. Sathiesh Kumar, and S. Anubha Pearline(✉)

Department of Electronics Engineering, Madras Institute of Technology, Anna University, Chennai 600044, India
vaitheeshwarir@gmail.com, sathieshkumar@annauniv.edu, anubhapearl@mitindia.edu

Abstract. Smart surveillance system is designed and developed to mitigate the occurrence of crime scenarios. Traditional image processing methods and deep learning approaches are used to identify the knife from camera feed. On identification of knife, the identity of person holding the knife is obtained using SSD ResNet CNN model. Also, an awareness alarm is generated by the system to caution the people in the surroundings. Experimental investigation clearly shows that the method of fine-tuned Xception deep learning model based on Separable Convolutional Neural Network (SCNN) with Logistic Regression (LR) classifier resulted in highest accuracy of 97.91% and precision rate of 0.98. Face detection is employed using a conditional face detection model based on SSD ResNet. The result obtained using deep learning approach is high compared to that of traditional image processing method. Real time implementation result shows that the model effectively detects the knife and identifies the person holding knife.

Keywords: Knife detection · Face detection · Deep learning · Finetuning · Smart surveillance

1 Introduction

Safety measure is an important constraint of human being for living a peaceful life. It is better to prevent the crime action rather than analyzing the footages after the crime incident. As per the statistics report by National Crime Records Bureau (NCRB-2016 and 2017), India, out of all violent crimes, murder occupies 7.1% and kidnapping about 20.5% of total population in which women are being highly targeted compared to men [1].

Mostly these violent crime involve knifes and firearms (guns) to threaten the person. Also, several criminal attack happen in public places and crowded areas. These actions are recorded in surveillance cameras. Police investigation often assists the help of surveillance camera footages to identify the offender as well as the defender. There are several steps taken by the Government, Researchers and Innovators to provide safety solutions for humans. Devices such as, Foot Wear Chip and SHE (Society Harnessing Equipment) has been implemented and safety measure application such as,

N. Nain et al. (Eds.): CVIP 2019, CCIS 1148, pp. 319–330, 2020.
https://doi.org/10.1007/978-981-15-4018-9_29

Raksha- women safety alert, VithU:V Gumrah Initiative and Shake2Safety are incorporated. Recently, the Government of Tamil Nadu, India, has launched a new application called "Kavalan". This application tracks the location of the victim in real time. All these precautious measure directly or indirectly involves the person's attention who is having the device or application to trigger the system. On the other hand, object detection and identification techniques are rapidly increasing using deep learning approach for several applications.

Thus, this paper aims to create a warning system based on deep learning concept is used to minimize the occurrence of crime incident by identifying the knife from the video feed and generates the alert sound to mitigate the crime action.

2 Related Work

Numerous work are reported by the researchers to detect the object in camera for safety purpose. Grega et al. [2] proposed an algorithm that is able to alert the human operator when a firearm or knife is visible in the image [2]. The authors implemented MPEG-7 feature extractor with Support Vector Machine (SVM) classifier and Canny edge detection, with MPEG-7 classifier for knife and firearm detection respectively.

Buckchash et al. [3] proposed a robust object detection algorithm. This proposed approach has three stages, foreground segmentation, Features from Accelerated Segment Test (FAST) based prominent feature detection for image localization and Multi-Resolution Analysis (MRA) [3]. The authors utilized Support Vector Machine (SVM) classifier for image classification and target confirmation. This method achieved about 96% accuracy in detecting the object.

Kibria et al. [4] proposed a comparative analysis of various methods for object detection, it involves HOG-SVM (Histogram of Oriented Gradients- Support Vector Machine), CNN (Convolutional Neural Network), pre-trained AlexNet CNN and the deep learning CNN methods are analyzed to detect object in the images. Authors reported that among all those methods CNN achieved the highest accuracy in detecting objects.

Yuenyong et al. [5] trained a deep neural network is on natural image (GoogleNet dataset) and fine-tuned to classify the IR images as person, or person carrying hidden knife [5]. By fine-tuning the GoogleNet trained on ImageNet dataset achieved 97% accuracy in predicting its classes.

Mahajan et al. [6] proposed a rescue solution for the safety of women as a wearable device-using microcontroller. The wearable device involves switch to trigger the shock circuit. An on-body camera and audio recorder is used to store the data in a SD card attached to the device. A GPS module is attached with the microcontroller device to track the location.

Harikiran et al. [7] proposed a security solution for women. The authors used a microcontroller based smart band and it is connected to a smart phone. The smart band proposed in the work consists of the several sensors to monitor the status of human and sends intimation to registered phone number in case of emergency.

From the literature, it is observed that researchers concentrated on detecting knives as a precautionary measure for ensuring people safety. So far, identification of the person holding the knife has not been carried out. Also, the researchers have not used conventional image processing techniques such as segmentation and windowing methods to identify objects in CCTV cameras. The observed sensitivity rate in the existing methods is less. Safety measures reported in the literature review resulted as wearable device. Therefore, any damages to the device disqualify the reliability of that device.

Hence, in the proposed a system is designed using deep learning neural network approach. The system automatically detects the dangerous objects such as knife in CCTV images and alerts about the hazardous situation with improved accuracy and precision rate. The proposed framework detects the knife using fine-tuned Xception deep learning model and identifies the person involved in the crime through face detection algorithm (SSD-Resnet CNN) from CCTV footages.

3 Methodology

Knife detection is employed by performing comparative analysis of various traditional image processing method and Xception deep learning model. After analyzing the various algorithms for the sensitivity and accuracy, the model with highest sensitivity value is selected for the real time implementation. The face identification system is implemented by using the conditional face detection algorithm based on SSD ResNet model.

The overall system workflow is shown in Fig. 1. The work involves using two different approaches, namely, traditional image processing and deep learning. In traditional image processing method, feature extraction is carried out using Local Binary Pattern (LBP), Haralick feature and Histogram of Oriented Gradients (HOG). The extracted features are classified using Machine Learning (ML) classifiers such as Support Vector Machines (SVM), Logistic Regression (LR) and Random Forest (RF). In deep learning method, Xception CNN model is utilized as feature extractor and classifier.

3.1 Traditional Image Processing Approach

Traditional image processing method involves extraction of features from the input images and classifying the images using ML classifier. At first, the input image is fed into the preprocessing unit. Then, the preprocessed image is fed to feature extraction block. After extraction of features from the images, it is flattened into a 1D array. The combined 1D feature vector Haralick-LBP feature and HOG features of the images in dataset are fed into three machine learning classifiers individually.

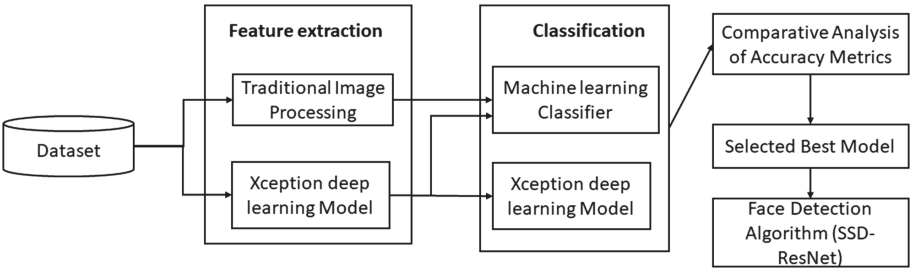

Fig. 1. Workflow for human safeguard measure.

Feature Extraction. To detect the knife in images, it is essential to extract its features such as texture, corner, edges etc. In this approach, the description of images in terms of its features is performed using three different methods.

Local Binary Pattern (LBP). It is a texture descriptor of an image introduced by Ojala [8]. For detecting knife, it is highly essential to find out its texture. Hence, the LBP pattern analysis is used. The standard LBP involves thresholding the center pixel (g_c) of 3×3 gray level matrix with its neighbor gray level intensities (g_i). LBP thresholding is given in Eq. (1). If g_i is lesser than g_c, the binary result of the pixel is set to 0 otherwise it is set to 1 [9].

$$LBP = \sum_{p=0}^{p-1} s(g_i - g_c)2^i \tag{1}$$

Where, 'p' is the number of neighbourhood points and 'i' is the neighbour pixel position.

Haralick Feature. Haralick feature is a texture feature extractor of an image introduced by Haralick [10]. It is obtained from Gray Level Co-occurrence Matrix (GLCM) of the image. GLCM computes the relationship between the intensities of neighboring pixel values [11]. GLCM is used to find the region of interest of an image by computing its gray level pixel intensities.

Histogram of Oriented Gradients (HOG). HOG descriptor is a feature extraction method that detects corners and edges of the object in the image. It is achieved using extraction of HOG feature [12]. The steps involved in computation of HOG are as follows.

Algorithm 1: HOG Feature Extraction
Input : Image, I **Output :** Feature vector (f_1, $f_{2,...}$ f_N)
Step 1: Normalization of Image – square root of color channels Step 2: Calculation of Gradients –Computing contour and silhouette Step 3: Cell formation and histogram computation Step 4: Normalizing local cell blocks (HOG feature values) Step 5: Converts all the hog descriptor for all blocks and combine as a HOG feature vector.

Machine Learning Classifiers

Logistic Regression. Logistic regression (LR) is a supervised binary classifier. Its performance is similar to SVM with linear kernel. It uses the logistic function (sigmoid function) to determine the probability of the predicted class. It predicts the probability by using Eq. (2) [13],

$$y = e^{b0+b1*x}/\left(1+e^{b0+b1*x}\right) \tag{2}$$

where, y is the predicted output, b0 is the bias or intercept term, b1 is the coefficient for the single input value (x) [13] and e is the Euler's value.

Support Vector Machine. Support Vector Machine (SVM) [14, 15] is a supervised learning algorithm, for classifying binary classes. SVM classifier is accomplished using Radial Basis Function (RBF) kernel. Linear kernel considers the hyperplane as a line. While the RBF kernel is based on Eq. (3) for creating the hyperplane to separate the classes [16]. In Eq. (3), $x^{(e)}$ and $x^{(k)}$ represent the feature value of class empty and class knife, respectively. γ is the boundary decision region.

$$K(x^{(e)}, x^{(k)}) = \exp\left(-\gamma\left|\left|x^{(e)} - x^{(k)}\right|\right|\right)^2, \gamma > 0 \tag{3}$$

Random Forest. Random forest (RF) is based on decision tree algorithm. It is an ensemble algorithm utilizing two or more methods for predicting the class [17]. The number of trees used in the work is 500. Random forest generates random subsets of tree, and aggregates the votes from the nodes for best selected feature values. It then averages the votes and the highest voted feature value class is considered as destination class [17].

3.2 Deep Learning Approach

Xception Deep Learning Model as Feature Extractor and Classifier. This is the second approach used in the studies. The performance of the traditional model with ML classifier resulted in lowest accuracy. In order to improve its accuracy a powerful deep learning model called 'Xception' is fine-tuned for feature extraction and classification. In this approach, the analysis is carried out in two different ways. Previous layer trainable parameter is set as a false or true. Setting Layer-trainable as 'false' considers the pre-trained weights from the Xception model and trains only the last three fine-tuned layers. On the other hand, setting layer-trainable as 'true' involves training the model from scratch.

Xception Deep Learning Model with ML Classifier. In this approach, the Xception pre-trained model is used for feature extraction. The extracted features are flattened to 1D vector and classified using ML classifiers such as random forest and logistic regression.

Fine-Tuned Xception Deep Learning Model. Xception model is the Extreme version of the Inception model. There are about 36 convolutional layers in Xception model followed by one fully connected layer, Global Average Pooling (GAP) and one output layer predicts the classes (Knife or Empty) using sigmoid activation function. The sigmoid function maps the feature values in the range between 0 and 1 [18].

It contains several depthwise separable convolution. This depthwise separable convolution is channel-wise nxn spatial convolution [18] and is followed by pointwise convolution. The mathematical representations of the convolution and depthwise separable convolutions are represented in Eqs. (4), (5) and (6) [19].

$$Conv(W,y)_{(i,j)} = \sum_{k,l,m}^{K,L,M} W_{(k,l,m)} y_{(i+k,j+l,m)} \tag{4}$$

$$PointwiseConv(W,y)_{(i,j)} = \sum_{k,l,m}^{K,L,M} W_{(m)} y_{(i,j,m)} \tag{5}$$

$$DepthwiseConv(W,y)_{(i,j)} = \sum_{k,l}^{K,L} W_{(k,l)} y_{(i+k,j+l)} \tag{6}$$

where, W is the weight matrix, y(i, j) is the image pixel coefficient, k, l, m is width, height and channel of the image, respectively.

3.3 Face Detection Algorithm

Once the knife in image is detected, the person holding the knife has to be identified. In the proposed work, a conditional face detection algorithm is used to identify the person holding knife. Hence, to incorporate the face detection method, SSD (Single Shot Detector) model is implemented. SSD is used for object detection. It is the fastest known model since it eliminates the need for region proposal of the object [20, 21]. SSD performs two operations. One is extracting the feature values and another one is to detect the object based on the convolution filter [21]. The architecture of face detection algorithm is shown in Fig. 2.

The steps involved in Conditional face detection algorithm are described in Algorithm 1. By implementing this algorithm, the model is effective in predicting the face of the attacker. This is due to conditional approach of the probability rate produced by the classifier. By doing so, only the faces in image where, the knife is detected in attacking position is highlighted with a bounding box.

Algorithm 1: Conditional Face detection

Input: Knife detected image(k), probability of the image p(k)
Output: Assaulter identified image

Step 1: Extract feature from the input image using Xception model
Step 2: Inspect the image for knife
Step 3: Once knife in image is detected, check its probability rate.

If (P (k) > 0.85)

i. Detect the faces in image.
ii. Implement the bounding box, to highlight the face of the assaulter. If the assaulter face is not clear, detect all the faces that are clearly identified as face by the model.

Else if (P (k)<0.85)

i. The face detection is not implemented
ii. Print only the probability occurrence of the knife and display the label as knife.

Step 4: Repeat step 1 to 3 for all the consecutive frames.

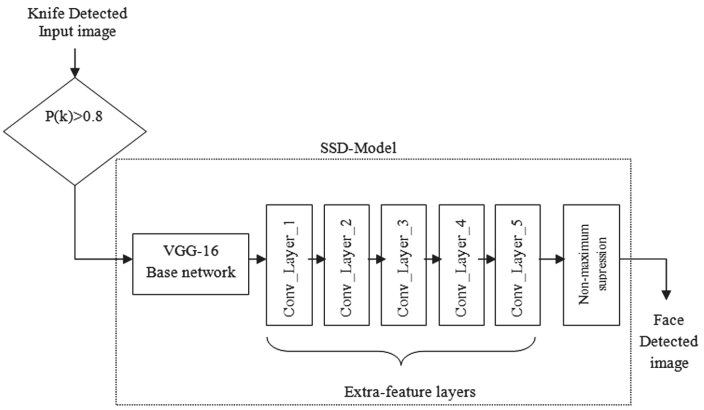

Fig. 2. Block diagram for conditional face detection algorithm using SSD-model

3.4 Dataset Description

The knife dataset is collected from two sources. The first source is Katedra Telekomunikacji [22], a university in Poland and another one is the datasets created by the students of Department of Computer Science, IIT Roorkee [23]. In addition to that, custom created real- time dataset is appended. One class consists of images with knife and another class consists of images without knife i.e., empty hand images. Thus, the three different dataset sources are collectively named as Dataset-weapon and used in the studies. The sample images from Dataset-weapon are shown in Fig. 3. The number of images in the datasets before and after augmentation is listed in Table 1. The analysis is carried for augmented dataset with train-test split ratio of 7:3 is considered.

Table 1. Dataset-weapon description

Classes	Number of images	
	Without augmentation	With augmentation
Positive (with knife)	3753	10891
Negative(without knife)	9750	10891
Total	13503	21782

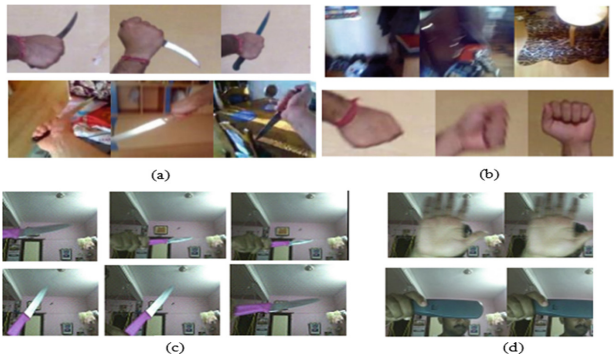

Fig. 3. Sample dataset (a) Positive images from database. (b) Negative images from database. (c) Positive images from real time dataset. (d) Negative image from real time dataset.

4 Results and Discussions

The reliability of the model depends on the performance metrics such as sensitivity, specificity and accuracy. The performance metrics considered in this paper are precision, recall, f1-score and accuracy.

4.1 Results of Traditional Image Processing Approach

The resulted performance metrics for the traditional feature extractors, Haralick and LBP with various ML classifiers are tabulated in Table 2. It is observed from Table 2, that the random forest classifier resulted in the highest accuracy of 86.7% compared to other classifiers.

Table 2. Accuracy metrics of Haralick-LBP features with ML classifiers

Classifier	Precision	Recall	F1 score	Accuracy (%)	Loss
SVM	0.74	0.74	0.73	74.02	0.45
LR	0.72	0.75	0.73	72.13	0.35
RF	**0.87**	**0.86**	**0.86**	**86.70**	**0.27**

Table 3. Accuracy metric of HOG features with ML classifiers

Classifier	Precision	Recall	F1-score	Accuracy (%)	Loss
SVM	0.65	0.67	0.65	68.08	0.49
LR	0.66	0.66	0.66	64.97	0.44
RF	**0.82**	**0.81**	**0.81**	**80.12**	**0.33**

Similarly, for HOG features, the performance metrics are tabulated in Table 3. From the Table 3, it observed that RF classifier resulted in highest accuracy with increased precision rate when compared to other classifiers. A good model should not only have highest accuracy but it must have highest sensitivity (precision) rate. Thus, precision value reveals how accurate the model is, while detecting the knife in images. Though considerable precision is achieved in Tables 2 and 3, the precision value is not sufficient to develop a precise prototype for real time implementation. Hence, the fine-tuned deep learning model is considered for both feature extraction and classification.

4.2 Results of Fine-Tuned Xception Deep Learning Model

In this approach, Fine-tuned Xception deep learning model is used as both feature extractor and classifier. The analysis is carried out using binary cross-entropy as loss function and ReLU as an activation function for 50 epochs. Adam is used as an optimizer with the learning rate of 0.001. The result of this approach is tabulated in Table 4.

Table 4. Analysis of Xception deep learning model

Method	Accuracy	Loss	Computation time	Number of parameters
Layer trainable false	86%	0.3	215 s/Epoch	54,528
Layer trainable true	**99.9%**	**0.02**	510 s/Epoch	2,29,07,128

It is observed that making layer trainable as 'true' attained global minima with increased accuracy of 99.9% and reduced loss of 0.02. The accuracy and loss plot of the model with respect to epochs is shown in Fig. 4(a) and (b).

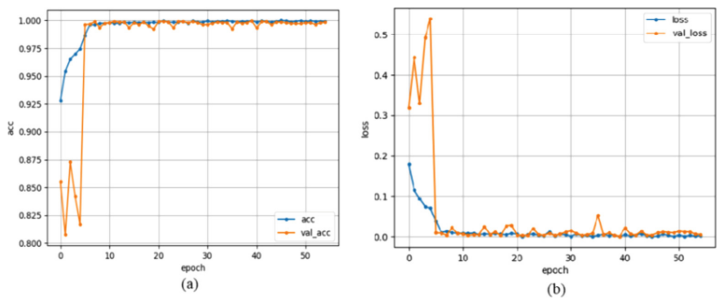

Fig. 4. (a) Accuracy plot (b) Loss plot of the Fine tuned Xception model

It is noticed from Fig. 4(a), the validating accuracy suddenly increasing from 4th epoch. This is due to the fact of optimization landscape. Hence, the activation function and optimizer attains global maxima in fourth epoch. Similarly, in Fig. 4(b), loss plot measures the inconsistency between the predicted outcomes with the true class. It is

observed from the plot that the model generates very low loss value of about 0.01. Precision and recall values for the fine-tuned model are 0.56. This is due to the output sigmoid layer is sensitive to the texture feature and falsely detects knife in the image.

4.3 Results of Fine-Tuned Xception Model with ML Classifier

In this approach, the last output layer with sigmoid activation function is replaced using random forest and logistic regression classifier. From Table 5, it is observed from the table that for LR classifier, the precision rate is 0.98 that indicates the highest reliability of the model. This is because of logistic regression that is meant for speeded confluence due to being zero-centered. The accuracy gained by the model is 97.84%.

Table 5. Performance metrics of Xception model with LR classifier and RF classifier

Class	LR classifier			RF classifier		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Empty	**0.97**	0.98	0.98	0.92	0.96	0.94
Knife	**0.98**	0.97	0.98	0.96	0.92	0.94

4.4 Real Time Implementation

From the analysis, it is observed that fine-tuned Xception model with LR classifier resulted in the highest accuracy and precision rate. Thus, it is selected for real time testing. The block diagram involved in real time prediction is shown in Fig. 5. It is carried out using Raspberry pi-3 board that consists of 4xARM cortex A53 processor with RAM about 1 GB.

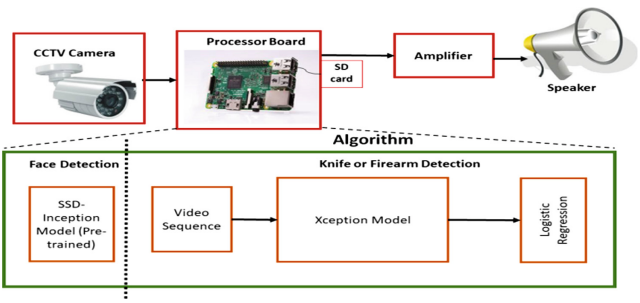

Fig. 5. Real time prediction of the CCTV frame.

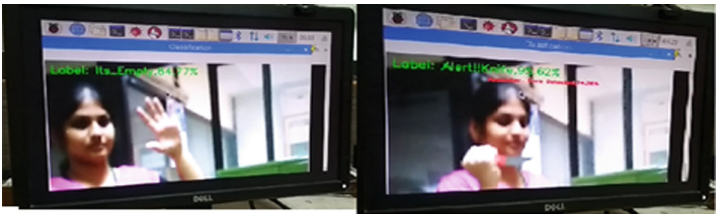

Fig. 6. Experimental setup of real time knife detection using Raspberry pi-3 board.

The real time implementation result is shown in Fig. 6. The camera lively monitors the environment. Once the knife is detected, the faces in the images are identified and the pre-recorded police siren sound start to alert the environment.

5 Conclusion

Knife detection in public places is an effective safety measure for human. The proposed method detects both knife and the person holding the knife (probably the offender). The work achieves highest accuracy of 97.84% utilizing the fine-tuned Xception model with LR classifier. The sensitivity rate achieved is 0.98. The prediction time of the model for real time data is less than a second. As a future work, the model will be implemented in real-time with increased number of images for different classes like gun, sword and other sharp objects.

Acknowledgement. The authors would like to thank NVIDIA for providing NVIDIA TITAN X GPU under University Research Programme.

References

1. Crimes in India-2016 Statistics - National crime records Bureau-Ministry of Home Affairs. https://timesofindia.indiatimes.com/realtime/Crime_in_India_2016_Complete_PDF.PDF
2. Grega, M., Matiolański, A., Guzik, P., Leszczuk, M.: Automated detection of firearms and knives in a CCTV image. Sensors **16**(1), 47 (2016)
3. Buckchash, H., Balasubramanian, R.: A robust object detector: application to detection of visual knives. In: IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 633–638. Hong Kong, China (2017)
4. Kibria, S.B., Hasan, M.S.: An analysis of feature extraction and classification algorithms for dangerous object detection. In: 2nd International Conference on Electrical & Electronic Engineering (ICEEE), December, pp. 1–4 (2017)
5. Yuenyong, S., Hnoohom, N., Wongpatikaseree, K.: Automatic detection of knives in infrared images. In: 2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), pp. 65–68 (2018)
6. Mahajan, M., Reddy, K.T.V., Rajput, M.: Design and implementation of a rescue system for safety of women. In: 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 1955–1959 (2016)

7. Harikiran, G.C., Menasinkai, K., Shirol, S.: Smart security solution for women based on Internet Of Things (IOT). In: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 3551–3554 (2016)
8. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
9. Meena, K., Suruliandi, A. Local binary patterns and its variants for face recognition. In: 2011 International Conference on Recent Trends in Information Technology (ICRTIT), pp. 782–786 (2011)
10. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **6**, 610–621 (1973)
11. Salhi, K., Jaara, E.M., Alaoui, M.T., Alaoui, Y.T.: GPU implementation of Haralick texture features extraction algorithm for a neuro-morphological texture image segmentation approach. In: 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), pp. 1–4 (2018)
12. Zhang, S., Wang, X.: Human detection and object tracking based on Histograms of Oriented Gradients. In: 2013 Ninth International Conference on Natural Computation (ICNC), pp. 1349–1353 (2013)
13. Dreiseitl, S., Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review. J. Biomed. Inform. **35**(5–6), 352–359 (2002)
14. Xiong, S.W., Liu, H.B., Niu, X.X.: Fuzzy support vector machines based on FCM clustering. In: 2005 International Conference on Machine Learning and Cybernetics, vol. 5, pp. 2608–2613 (2005)
15. Marsland, S.: Machine Learning: An Algorithmic Perspective. Chapman and Hall/CRC, Boca Raton (2011)
16. Non-linear SVM classification with kernels (2011). https://www.google.com/url?q=http://openclassroom.stanford.edu/MainFolder/DocumentPage.php?course%3DMachineLearning%26doc%3Dexercises/ex8/ex8.html
17. Liaw, A., Wiener, M.: Classification and regression by random forest. R News **2**(3), 18–22 (2002)
18. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition, pp. 1251–1258 (2017)
19. Kaiser, L., Gomez, A.N., Chollet, F.: Depthwise separable convolutions for neural machine translation (2017). arXiv preprint. arXiv:1706.03059
20. Karpathy, A.: CS231n: Convolutional Neural Networks for Visual Recognition. http://cs231n.github.io/convolutional-networks
21. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
22. Knives images database. http://kt.agh.edu.pl/∼matiolanski/KnivesImagesDatabase/
23. Knives Dataset. https://www.sites.google.com/site/kdsdataset/