Computers, Materials & Continua
DOI: 10.32604/cmc.2022.020107
Article

Tech Science Press

Efficient Deep Learning Modalities for Object Detection from Infrared Images

Naglaa F. Soliman[1,2], E. A. Alabdulkreem[3], Abeer D. Algarni[1,*], Ghada M. El Banby[4], Fathi E. Abd El-Samie[1,5] and Ahmed Sedik[6]

[1]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, 84428, Saudi Arabia
[2]Department of Electronics and Communications, Faculty of Engineering, Zagazig University, Zagazig, 44519, Egypt
[3]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, 84428, Saudi Arabia
[4]Department of Industrial Electronics and Control Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, 32952, Egypt
[5]Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menouf, 32952, Egypt
[6]Department of the Robotics and Intelligent Machines, Faculty of Artificial Intelligence, KafrelSheikh University, Kafr el-sheikh, Egypt
*Corresponding Author: Abeer D. Algarni. Email: adalqarni@pnu.edu.sa
Received: 09 May 2021; Accepted: 21 December 2021

Abstract: For military warfare purposes, it is necessary to identify the type of a certain weapon through video stream tracking based on infrared (IR) video frames. Computer vision is a visual search trend that is used to identify objects in images or video frames. For military applications, drones take a main role in surveillance tasks, but they cannot be confident for long-time missions. So, there is a need for such a system, which provides a continuous surveillance task to support the drone mission. Such a system can be called a Hybrid Surveillance System (HSS). This system is based on a distributed network of wireless sensors for continuous surveillance. In addition, it includes one or more drones to make short-time missions, if the sensors detect a suspicious event. This paper presents a digital solution to identify certain types of concealed weapons in surveillance applications based on Convolutional Neural Networks (CNNs) and Convolutional Long Short-Term Memory (ConvLSTM). Based on initial results, the importance of video frame enhancement is obvious to improve the visibility of objects in video streams. The accuracy of the proposed methods reach 99%, which reflects the effectiveness of the presented solution. In addition, the experimental results prove that the proposed methods provide superior performance compared to traditional ones.

Keywords: Deep learning; object detection; military applications; OFDM; SPIHT; IoT

1 Introduction

Identifying objects is the most critical issue in different applications such as military, medical, and industrial applications. Generally, surveillance systems depend on the idea of using certain types of cameras to get video streams in order to track, detect and identify objects, easily. For military applications, it is necessary to determine the threats represented in weapons or guns in order to take the necessary precautions in emergency situations. Therefore, there is a bad need to design automatic recognition systems that are able to detect weapons, easily. Multi-spectral imaging techniques are widely used in surveillance systems. These systems include microwave imaging systems as well as visual light imaging systems. In addition, thermal IR cameras are good candidates for surveillance applications.

There are several different techniques to perform scene imaging as IR imaging, microwave imaging and visual imaging. Visible light cameras exploit the reflected light from objects to form the desired image. Thermal images acquired with IR cameras depend on temperature variations emitted from the body followed by a focusing process by the camera optics on the IR detector. Complex algorithms are used to translate the data that comes from the IR detector into an image [1]. Microwave imaging systems can be used for detecting the concealed objects, but this technique has a limited use as the microwave frequencies reveal all body details in contradiction with privacy concerns [2]. All imaging systems should perform several internal processing in order to get the clear crisp image. Basically, the beams from the light source strike the imaging sensor, and then the produced electrical charge will be converted into a digital signal representing the intensity of the incident light on the imaging sensor. Quantization and filtering are applied as digital pre-processing steps to generate the required digital image. The most widely-used imaging system for night vision is the thermal imaging system. This mechanism depends on thermal energy emission from the objects due to the temperature variations. With this type of imaging, concealed weapons will appear darker with low intensity due to the variations of temperature between the person's body and the weapon object. Imaging based on thermal IR radiation is called thermography, in which the electronic capturing device detects the thermal energy patterns due to variations of temperature of objects. Feature extraction is the most powerful tool to classify and identify unknown concealed objects. In the proposed approach, the concept of deep neural network is used to generate distinctive feature maps that can be used to determine the type of the concealed weapon.

Recently, wireless communication has been widespread in several applications such as 5G networks, Wireless Sensor Networks (WSNs), Internet of Things (IoT), and health monitoring. The Orthogonal Frequency-Division Multiplexing (OFDM) has been considered one of the state-of-the-art solutions that are able to fulfill the requirements of the next generation communication systems, such as the increased capacity of users and the mitigation of multipath fading. The OFDM is employed in the physical layer of broadband wireless networks to allow high data rate communication for both civilian and military applications. It is used as a technology to allow superior spectral efficiency, low cost and resistance to Inter-Symbol Interference (ISI) effects [3–5]. In addition, compression algorithms are badly needed in IoT applications in order to reduce the size of the multimedia content. Furthermore, in the Internet environment, to allow interactive viewing more efficiently, compression schemes are required. Compression should be inherently scalable and it should support a high degree of random accessibility. Set Partitioned in Heretical Tree (SPHIT) is a lossless and reversible compression technique that can be used to allow complete recovery of the original image. It is scalable in nature, with low complexity, and a proficient encoding process. The SPIHT was firstly introduced by Said et al. [6] based on the wavelet transform. It has a better objective performance and less block artifacts compared with other compression techniques that are based on the DCT such as the JPEG

[7,8]. The encoding is performed by quantizing and coding of the wavelet coefficients to select the significant coefficients to be used as the roots. The decomposition structure is assumed to be the octave-band structure, and then the sub-bands at different levels are used with some sort of alignment.

Different studies have utilized the SPIHT algorithm for image transmission over wireless systems [9,10], because it has a good rate-distortion performance for still images with comparatively low complexity, and it is scalable or completely embeddable. The success of this algorithm in compression of images with high efficiency and simplicity makes it well known as a benchmark for embedded wavelet image coding [11,12]. In this framework, SPHIT is employed as a source coder tool for robust wireless video stream transmission, as it is a progressive coding algorithm. It is flexible in achieving the required code rate and also simple to design. The compression of images and videos for drone applications is important due to the high temporal and spatial correlation between pixels leading to redundancies. Firstly, the image stream captured by drones is different from that transmitted over other networks due to the enormous data size, and power required for each node in processing and transmission. Then, compressed images are transmitted over the OFDM wireless channel to reduce the communication burden. At the receiver side, a deep learning model can be used for object detection as in the detection of concealed weapons.

Raturi et al. [13] studied the urgent need to develop automated monitoring systems. Their proposed security surveillance system has the ability to detect any type of concealed objects such as firearms or even any type of weapon including knife, and scissors, which may cause security problems. They introduced a framework to detect and classify the concealed weapons by analyzing the Closed-Circuit Television CCTV stream data. Their classification framework depends on deep-learning-based object detection and classification techniques. To detect the concealed weapons, a multi-sensor stream data capturing framework is used, and the concept of sensor fusion is adopted to generate integrated images. These integrated images are further processed for segmentation to localize objects of interest. The authors utilized an R-CNN (Region-based Convolutional Neural Network) model to classify weapons from the collected dataset. They achieved a 93.6% efficiency in the detection of concealed weapons. Kaur et al. [14] proposed a method to extract features from input images to detect and classify different shapes of guns based on Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) techniques. They achieved a 90% efficiency using both methods with faster processing for the SURF algorithm. Mahajan et al. [15] introduced a survey of the different methods used for the detection of concealed objects to improve the accuracy and effectiveness of the security system. They reviewed the presented systems that can be used for the recognition of different objects based on extracted features. In addition, they introduced motion analysis to detect concealed weapons from video frames.

Chadha [16] studied the subject of Concealed Weapon Detection (CWD), because it was of major interest in the United States. In this work, acoustic signals have been used to detect a concealed weapons. Acoustic signals have been selected for this task, as they are characterized by minimum health risks and maximum cost effectiveness. The reflected acoustic signals are used as signatures that depend on the object composition and shape. Acoustic signatures are compiled to identify the concealed weapons. Fast Fourier Transform (FFT) is used to extract features from each acoustic signature. A mathematical model is created through machine learning to detect concealed weapons. An accuracy of 88% was achieved based on Random Forest (RF) classifier algorithm. Hussin et al. [17] succeeded to detect the desired object based on Circular Hough Transform (CHT). Color pre-processing, filtering, smoothing and shape detection are the steps to be applied to detect and allocate the desired objects.

Deep learning has been extensively involved in several applications, such as medical [18], security [19] and biometric recognition [20]. Hence, this paper is concerned with deploying Deep Learning Models (DLMs) on both IR sequences and drone surveillance images in military applications. This paper presents an efficient model for weapon detection. The target of deploying deep learning is to recognize the type of the weapons carried by persons or soldiers. The images can be collected by distributed sensor networks and/or surveillance drones. The contributions of the proposed work can be illustrated as follows:

1- Proposal of a hybrid surveillance framework based on both wireless sensor networks and drones.
2- Compression of the obtained images using SPIHT encoder and then transmission of the compressed images via OFDM channel.
3- Building a CNN model which provides an optimal performance for weapon detection.
4- Building a hybrid model based on CNN and ConvLSTM models for weapon detection.
5- Providing a comparative study between the two proposed models from the accuracy of detection perspective.

This paper is structured into five sections. Section two represents the proposed framework. The simulation results are discussed in section three. Moreover, brief result discussion and comparison are represented in section four. Finally, section five shows the conclusions of this paper.

2 Proposed Framework

This paper presents a hybrid surveillance framework based on wireless sensor networks and drones for military applications. A hybrid surveillance strategy is introduced through the received images acquired using OFDM modulation. The motivation for this strategy is the lake of long-time missions for drones. The drone devices are supplied with batteries or solar cells, which cannot be sustainable for long-time surveillance missions. Hence, there is a need to investigate a hybrid system that includes a sensor network with low power consumption to serve the sustainable surveillance mission. The wireless sensor networks are launched to detect weapons in the surrounding sites. Then, if a weapon is detected, drones can be used to make short-time missions. Such missions could be either surveillance or defense missions. Fig. 1 shows the proposed framework, while Fig. 2 shows the block diagram of the proposed scheme for weapon detection.

The proposed scheme consists of multiple transmitters which are used within distributed wireless sensor networks. The sensors transmit images through wireless channels that are assumed to be OFDM channels. Finally, the detected images are received by a central receiver. This receiver is arranged to classify the detected weapons using deep learning. In the next section, the proposed DLMs are discussed in detail.

2.1 SPIHT-OFDM Model

The block diagram of SPIHT followed by OFDM modulator (SPHIT-OFDM) transmitter is shown in Fig. 3. Firstly, the input video stream is split into frames, and each frame is processed, separately. The coefficients of each frame are categorized into different layers based on the significance of the input stream. Afterwards, the modified stream of each frame is adapted to binary format that is appropriate to be processed by the OFDM modulator.

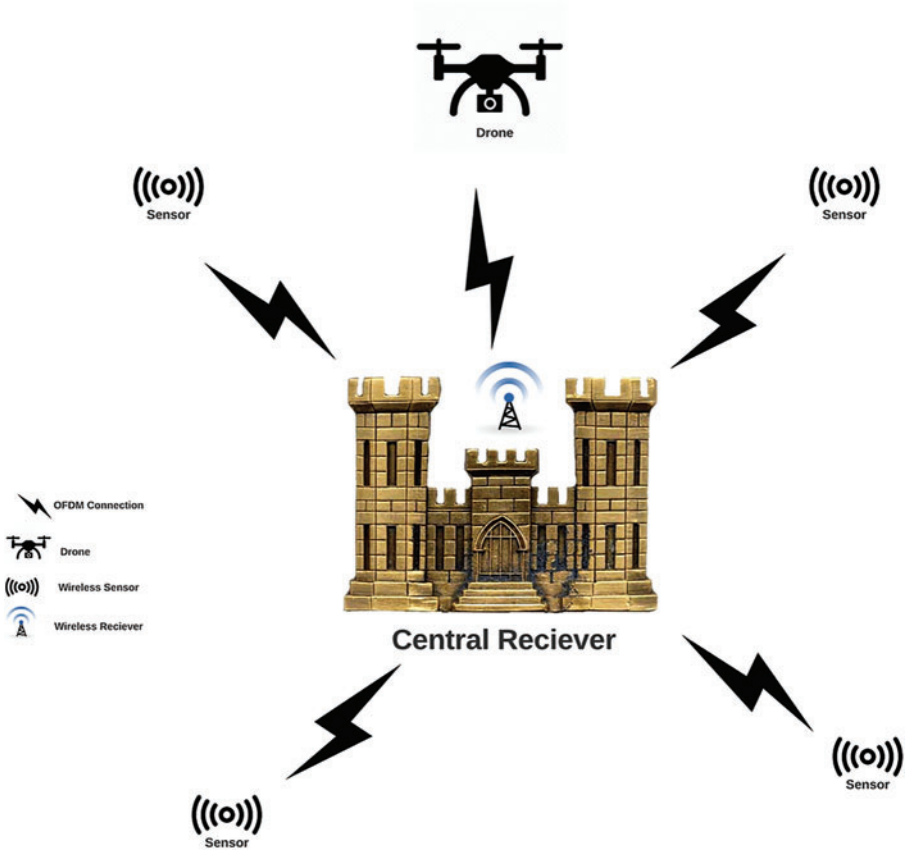

Figure 1: The proposed SDN military surveillance framework

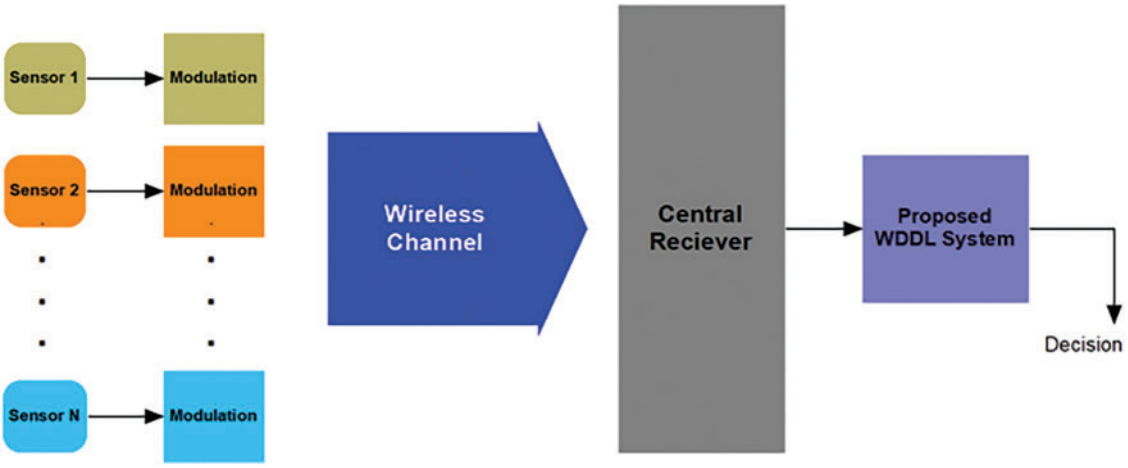

Figure 2: The proposed WDDL scheme

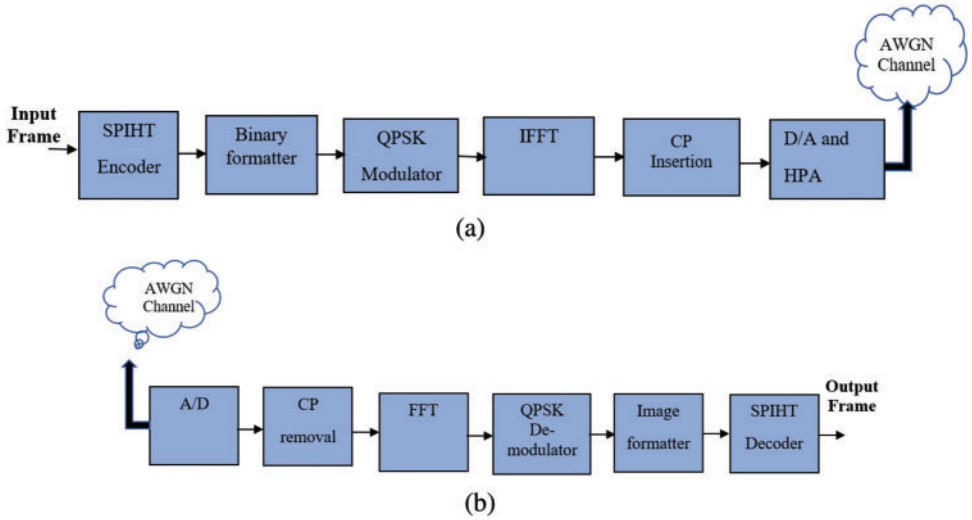

Figure 3: Block diagram of the SPIHT-OFDM transmitter and receiver. (a) SPIHT-OFDM transmitter, (b) SPIHT-OFDM receiver

In the OFDM modulator, the input bit stream is divided into several overlapped sub-channels, namely N sub-channels using the serial-to-parallel (S/P) converter. The duration of the new data symbol is smaller than that of the original one in order to mitigate the multi-path delay spread and provide enhancement of the spectral efficiency. In addition, the multi-path propagation problem is solved by converting the wideband frequency selective channels into N overlapped frequency selective channels. Hence, the bit streams on each sub-carrier are mapped into the data symbols (constellation points) by Quadrature-Phase Shift Keying (QPSK). Then, the bit stream is modulated by a set of orthogonal sub-carriers $\{f_k, k = 0, 1, \ldots, N-1\}$ using IFFT to construct the OFDM signal $X_{k,z} = \{X_k, k = 0, 1, \ldots, N-1\}$. Each data symbol is assigned to a different sub-carrier [21–23].

The complex envelope of the transmitted OFDM signal is represented by the following equation [22]:

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi f_k t}, 0 \leq t < T_s \tag{1}$$

where X_k represents the constellation points to be transmitted at the carrier frequency f_k, T_s is the data symbol period, and f_k is the center frequency of the k^{th} sub-carrier.

To avoid the problems of Inter-Symbol Interference (ISI) and Inter-Carrier Interference (ICI), a Cyclic Prefix (CP) is pre-appended to each data block of every sub-carrier. This introduces a cyclic redundancy of length N_g samples. The insertion the CP during the guard interval converts the linear convolution to a circular convolution. Finally, the signal samples $x(n)$ are embedded into the Digital-to-Analogue converter (D/A) and amplified using the High Power Amplifier (HPA). The OFDM signal $x(t)$ is transmitted over an Additive White Gaussian Noise (AWGN) channel.

At the receiver side, the received signal is given by:

$$r(t) = x(t) * h(t) + n_o(t) \tag{2}$$

where $r(t)$ is the received signal, $h(t)$ is the channel impulse response and $n_o(t)$ is a complex AWGN with single-sided power spectral density N_o.

At the receiver side, the received signal is converted to digital format using an (A/D) converter. Then, all processes are reversed.

2.2 Proposed WDDL System

In this paper, a Weapon Detection Deep Learning (WDDL) system is proposed. It consists of three phases. The first phase is the pre-processing phase. The second one is the forward propagation phase. Finally, the third one is the back-propagation phase. Fig. 4 shows the block diagram of the proposed deep learning model for the classification task.

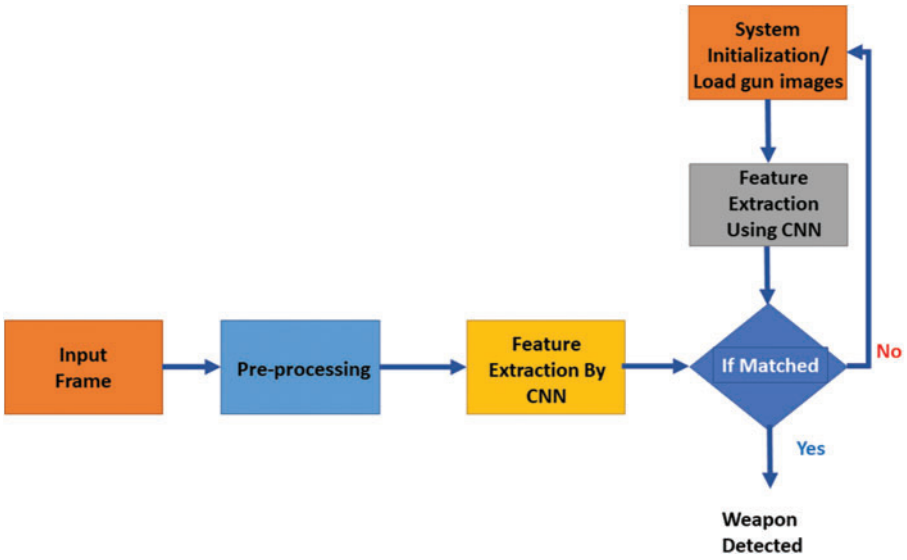

Figure 4: Block diagram of the classification stage of the proposed WDDL system

2.2.1 Video Stream Pre-Processing

In this section, the methodology to perform digital pre-processing for each frame of the video stream is explained. Frame pre-processing allows improvement of its quality prior to any detection task. Such improvement is accomplished by eliminating unwanted distortions and enhancing the important features, so that the CNN model can perform its classification task, efficiently. Adaptive Weiner filter is one of the most common techniques used to improve the visibility of digital images by suppressing noise and preserving edges. The operation of the adaptive Weiner filter depends on estimating the local mean and variance around each pixel to minimize the Mean Squared Error (MSE) as follows [24]:

$$MSE(\hat{x}) = \frac{1}{K}\sum_{i,j=1}^{K}(\hat{x}(i,j) - x(i,j))^2 \tag{3}$$

where $x(i,j)$ is the required noise-free image, $\hat{x}(i,j)$ is the estimated Weiner filtered image, and K is the total number of image pixels.

The Weiner filter update equation is given by:

$$\hat{x}(i,j) = \frac{\sigma_x^2(i,j)}{\sigma_x^2(i,j) + \sigma_n^2(i,j)}[y(i,j) - \mu_x(i,j)] + \mu_x(i,j) \tag{4}$$

where $y(i,j)$ is the noisy image, $\sigma_x^2(i,j)$ is the noisy image local variance, $\mu_x(i,j)$ is the noisy image local mean, and $\sigma_n^2(i,j)$ is the noise local variance.

For the local mean and local variance estimation, we get:

$$\widehat{\mu}_x(i,j) = \frac{1}{(2r+1)^2} \sum_{p=i-r}^{i+r} \sum_{q=j-r}^{j+r} y(p,q) \tag{5}$$

$$\sigma_x^2(i,j) = \frac{1}{(2r+1)^2} \sum_{p=i-r}^{i+r} \sum_{q=j-r}^{j+r} (y(p,q) - \widehat{\mu}_x(i,j))^2 - \sigma_n^2(i,j) \tag{6}$$

Moreover, histogram equalization is applied on the output image after Weiner filtering in order to enhance the image contrast, so that the features are extracted with the CNN. For L different gray levels, each level i occurs n_i times out of a total of n occurences. The Transform Function (T.F) based on histogram equalization to obtain an image with better contrast can be estimated with the following relation [25]:

$$T.F = \left(\frac{n_0 + n_1 + \ldots + n_i}{n} \right) (L-1) \tag{7}$$

2.2.2 Feature Extraction by CNN

In this subsection, the two proposed WDDL models based on CNN and ConvLSTM are introduced. These models are designed to extract the image features, efficiently. Hence, the effective design is obtained by setting hyper-parameters including number of layers, number of filters, filter size in each layer, number of epochs and batch size to certain values to enhance the efficiency of classification.

The CNN is considered an efficient feature extraction methodology. The main idea of the CNN is to perform the 2D digital filtering on the input images to extract the feature maps that contain distinctive features from the images [26–28]. These filters have two main control parameters: filter size and stride. The filter size represents the dimensions of the applied filters and the stride represents the position of the filter initial sliding point. Moreover, if the filters have a size of $M \times M$, the stride size would be $(M-1)/2$. The pooling layer is mainly used for feature reduction, where the feature map is segmented into pixel windows, and each window is reduced into one pixel. The pooling process is controlled by two parameters. The first one is the type of pooling, which is performed on the feature map, while the second is the window size. There are two types of pooling processes: max-pooling and average pooling. The max-pooling process elects the maximum value from each window, while the average pooling reduces the window pixels into their mean value.

To produce a distinctive feature map, the input image is inserted into the feature extraction network which consists of sequential pairs of convolutional and pooling layers. These features are involved in the classification process with a neural network to produce the output of the deep neural network. Through the training stage, the weights of all layers are regulated using the back-propagation algorithm.

The first proposed approach comprises three convolutional (CNV) layers, each trailed by a maximum pooling layer. A Global Average Pooling (GAP) layer is finally inserted. A final classification layer with two classes (weapon or no weapon) is utilized. The CNV layers work as feature extractors, where each CNV layer gives its corresponding feature map. Max-pooling after each CNV layer allows a resizing process of the feature map in order to keep only the most representative features. The output

of the last pooling layer is fed into the GAP layer. Tab. 1 shows a summary of the first proposed approach based on a CNN.

Table 1: Summary of the proposed DLM based on CNV layers

Layer type	Output shape
Batch normalization	(224, 224, 3)
Reshape to ConvLSTM	(1, 224, 224, 3)
ConvLSTM	(1, 224, 224, 16)
Reshape to Conv.	(224, 224, 16)
Conv.	(222, 222, 32)
Pooling	(111, 111, 32)
Conv.	(109, 109, 64)
Pooling	(54, 54, 64)
Conv.	(52, 52, 128)
Pooling	(26, 26, 128)
Global average layer	(128)
Dense	(2)

In the second proposed approach for object detection from IR image sequences, a ConvLSTM layer and three CNV layers followed by max-pooling layers are used. The ConvLSTM layer has an advantage of remembering the previous state of extracted features and constructing a series of feature states in some sense of state prediction based on the state history. Unfortunately, this characteristic may be undesired. Assuming that a certain state in the series is dropped, all next states would be affected and dropped, also. Hence, the design of such a model should be performed carefully to be reliable and sustainable in the testing phase. Tab. 2 presents the hyper-prameters of the second proposed hybrid model based on a ConvLSTM layer and CNV layers.

Table 2: Summery of the second proposed DLM based on a hybrid structure of a ConvLSTM layer and CNV layers

Layer type	Output shape
Conv.	(222, 222, 16)
Pooling	(111, 111, 16)
Conv.	(109, 109, 32)
Pooling	(54, 54, 32)
Conv.	(52, 52, 64)
Pooling	(26, 26, 64)
Conv.	(24, 24, 128)
Pooling	(12, 12, 128)
Conv.	(10, 10, 256)
Pooling	(5, 5, 256)

(Continued)

Table 2: Continued

Layer type	Output shape
Global average layer	(256)
Dense	(2)

3 Simulation Results

3.1 Data Description

In this paper, the proposed DLMs are tested on the Terravic Weapon IR dataset [29]. This dataset is suitable for weapon detection and weapon discharge detection with thermal imagery. It has been collected using a special thermal sensor (Raytheon L-3 Thermal-Eye 2000AS). Terravic Weapon IR dataset consists of five thermal sequences. Each sequence includes IR frames for a certain weapon as shown in Tab. 3. The included frames are formatted into 8-bit gray-scale JPEG images. The size of each frame is 320 × 240 pixels. Samples of the dataset are presented in Fig. 5.

Table 3: Description of weapon IR dataset

Sequence no.	No. of frames	Description
#1	397	Subject walks with a holstered gun
#2	245	Video grouping demonstrating gag impact and shell release. Make: Colt Model – AR15 (6601 Match HBAR). Type: .223 (5.56 × 45). Limit: 30 round magazines utilized for recordings. Material: Steel w/manganese phosphate covering and aluminum w/hard anodized covering. Polymer buttstock, hold and forestock.
#3	480	Video succession demonstrating gag impact and shell release. SKS Yugoslavian made, Assault rifle Shoots 7.62 × 39, or 7.62 NATO. Appeared with discretionary 30 Rd. mag.
#4	481	Video succession indicating gag impact and shell release. Make: Smith and Wesson. Model: 5906. Type: 9 mm. Limit: 15 + 1. Material: Stainless steel.
#5	297	Video arrangement indicating gag impact and shell release. Make: Colt Model 150; AR15 (6601 Match HBAR). Type: .223 (5.56 × 45). Limit: 30 round magazines utilized for recordings. Material: Steel w/manganese phosphate covering and aluminum w/hard anodized covering. Polymer buttstock, grasp and forestock.

Figure 5: Samples of the dataset

3.2 Evaluation Metrics

In this proposed approach, accuracy is used to estimate the strength of the CNN model and give indication if there is a misleading result. Accuracy is calculated as follows:

$$Accuracy = \frac{Sum\ (true\ positive\ and\ true\ negative)}{Sum\ (total\ IR\ population)} \times 100 \tag{8}$$

3.3 Results

The simulation tests are performed on the dataset using Python 3.5, TensorFlow and Keras. The proposed models have been assessed, in the training stage, based on k-overlay cross validation. This strategy continues by arbitrarily fragmenting the dataset into k folds of roughly equivalent size. Out of these folds, (k-1) folds are used for training and the rest one is used for testing. In our simulation experiments, 90% of the data is used for training and the other 10% are used for testing. This process is repeated on the data through a folding process several times, and the results are averaged. Fig. 6 illustrates the data fragmentation process.

Figure 6: Fragmentation of the dataset

In the first simulation experiment, the proposed DLM is carried out on the IR images of soldiers. In the experiments, the number of epochs and batch size are set to 10 for each. The accuracy of the two proposed models is illustrated in Figs. 7 and 8. In addition, the loss of the two proposed models is illustrated Figs. 9 and 10. It can be observed that the accuracy in Figs. 7 and 8 is increased gradually till it gets to 100% at epoch No. 9. Although the IR images have low resolution, the proposed approach achieves an accuracy of 100% in robust conditions. Tab. 4 shows the value of accuracy and loss at each epoch for the first and the second proposed approaches.

In the second simulation experiment, the performance of the two proposed DLMs is investigated on the compressed images transmitted over OFDM channel. The proposed models are also carried out on the compressed video frames. The compression is performed due to the large amount of data between the distributed sensors and the central receiver. The compressed video stream is broadcasted over an OFDM channel. This experiment is carried out to assess the performance of the proposed DLMs in the wireless communication scenario. The accuracy of the two proposed models is illustrated in Figs. 11 and 12. In addition, the loss results are illustrated in Figs. 13 and 14. A comparison of the loss and accuracy between the proposed models on the compressed received images is introduced in Tab. 5. The simulation results reveal that the proposed DLMs achieved accuracies of 99.2%, 99.5% for CNN and ConvLSTM models, respectively, as shown in Tab. 5.

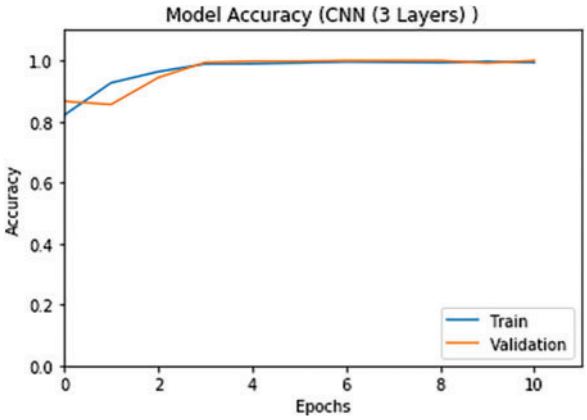

Figure 7: Accuracy of the 1st model

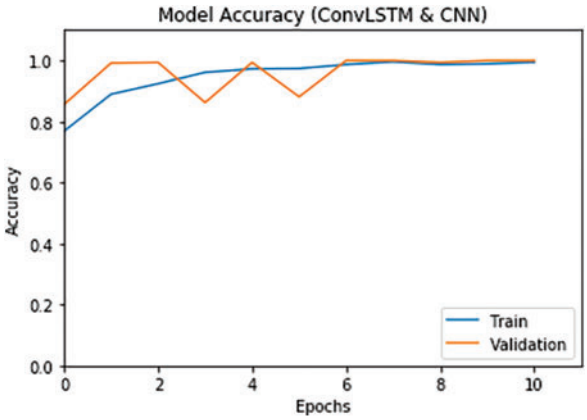

Figure 8: Accuracy of the 2nd model

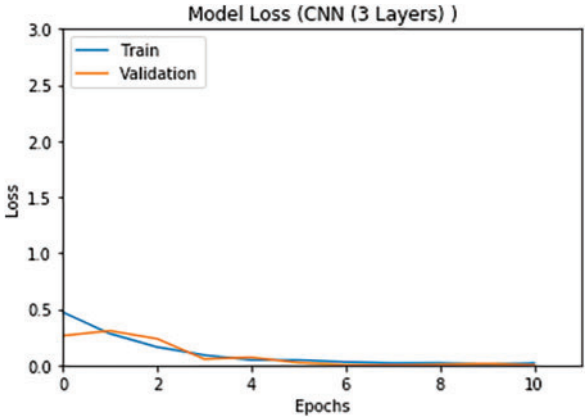

Figure 9: Loss of the 1st model

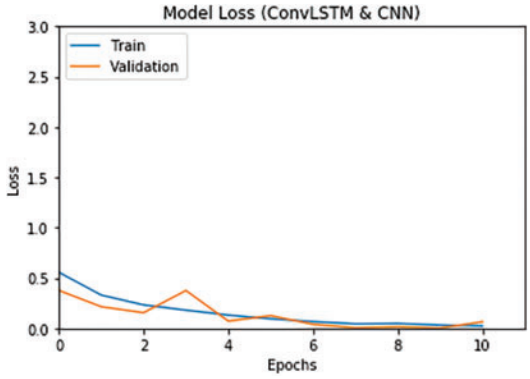

Figure 10: Loss of the 2nd model

Table 4: Accuracy and loss at each epoch in the first experiment

Epoch No.	CNN			ConvLSTM & CNN		
	Loss	Accuracy (%)	Testing time	Loss	Accuracy (%)	Testing time
1	0.357	85	1.04 Sec.	0.377	85	3.92 Sec.
2	0.264	86		0.255	90	
3	0.236	94		0.155	95	
4	0.1	96		0.375	86	
5	0.07	98		0.12	97	
6	0.05	98.5		0.07	98	
7	0.03	99		0.04	99	
8	0.01	99		0.005	99	
9	0.0029	100		0.0031	100	
10	0.0015	100		0.0012	100	

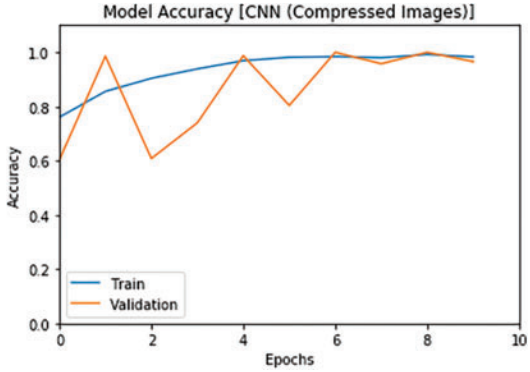

Figure 11: Accuracy of the 1st model

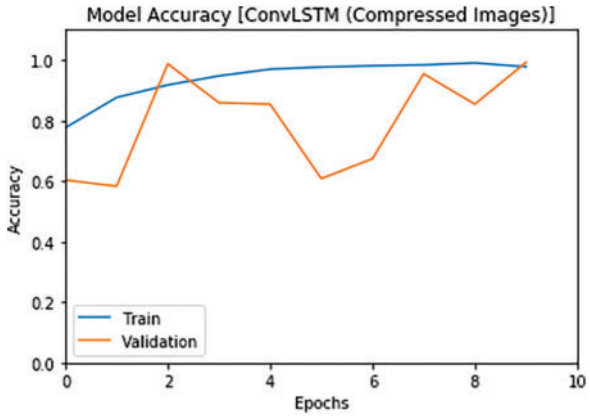

Figure 12: Accuracy of the 2nd model

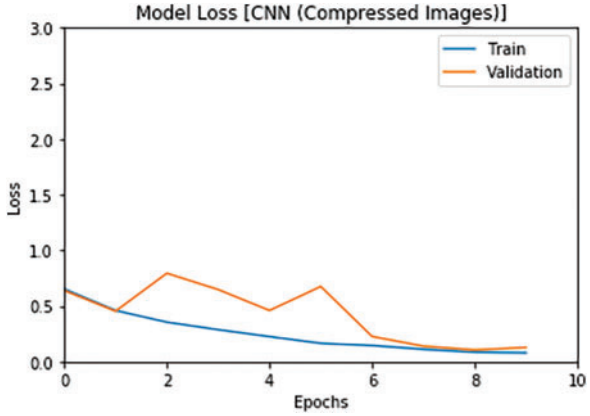

Figure 13: Loss of the 1st model

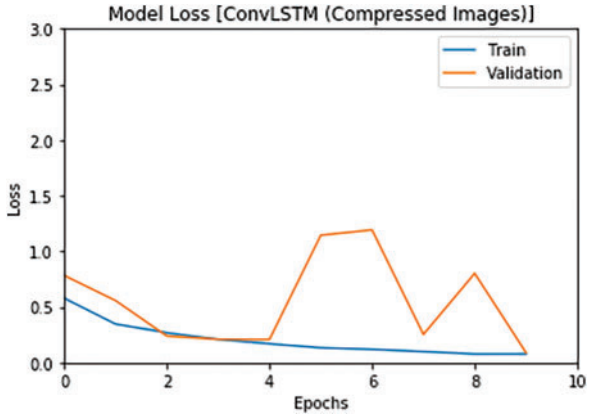

Figure 14: Loss of the 2nd model

Table 5: Accuracy and loss at each epoch in the second experiment

Epoch No.	CNN			ConvLSTM & CNN		
	Loss	Accuracy (%)	Testing time	Loss	Accuracy (%)	Testing time
1	0.631	60.3	1.87 Sec.	0.639	60.3	3.92 Sec.
2	0.453	90.6		0.557	61.2	
3	0.794	60.8		0.238	98.7	
4	0.459	86.9		0.209	86	
5	0.209	98.8		0.208	86.5	
6	0.675	86.8		0.145	90.8	
7	0.226	98		0.169	88.7	
8	0.139	100		0.252	96.1	
9	0.105	100		0.055	98.2	
10	0.109	99.2		0.008	99.5	

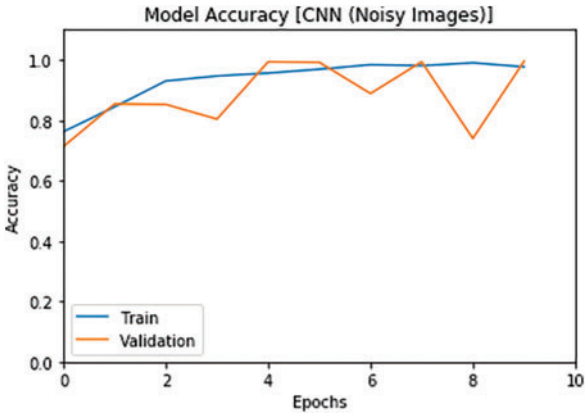

Figure 15: Accuracy of the 1st model

Finally, the performance of the proposed DLMs is evaluated on noisy image. In this case, the transmitted signal after OFDM is affected by the existence of AWGN. The AWGN is induced, because the distance between the distributed sensors and central receiver is assumed to be within the range of 1 km. The target is to detect any suspicious motion from such an intruder (soldier) and detect the type of the gun with him. So, the proposed DLMs are carried out on noisy images in order to perform this task. The accuracy and the loss results are introduced in Figs. 15–18 and Tab. 6. The simulation results ensure that the proposed DLMs achieve accuracies of 96.5%, and 99% for CNN and ConvLSTM, respectively.

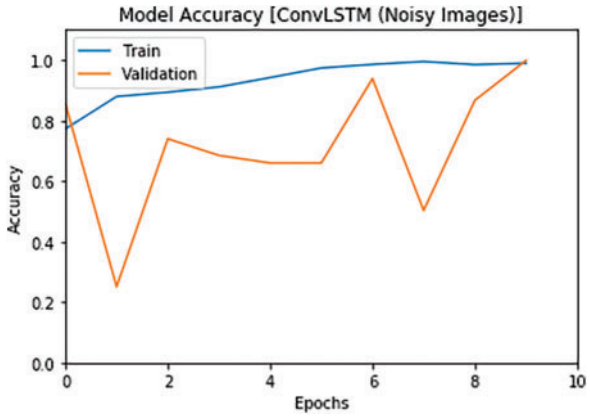

Figure 16: Accuracy of the 2nd model

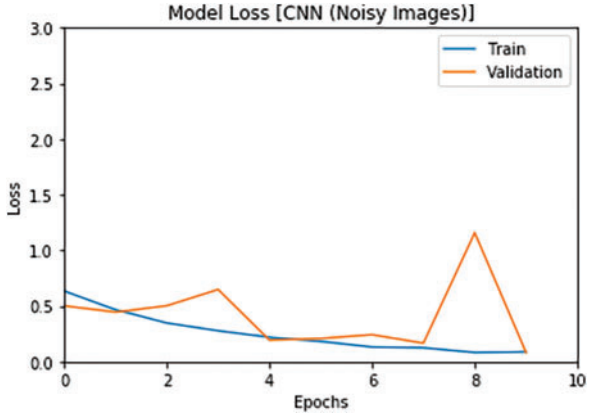

Figure 17: Loss of the 1st model

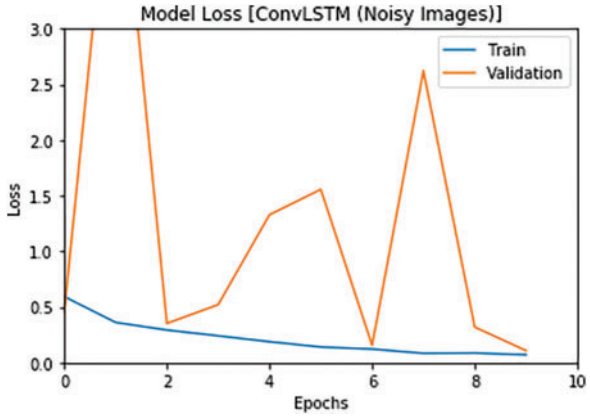

Figure 18: Loss of the 2nd model

Table 6: Accuracy and loss at each epoch in the third experiment

Epoch No.	CNN			ConvLSTM & CNN		
	Loss	Accuracy (%)	Testing time	Loss	Accuracy (%)	Testing time
1	0.701	76.6	1.87 Sec.	0.639	79.3	3.92 Sec.
2	0.462	90.6		0.557	88.2	
3	0.376	97.4		0.238	98.7	
4	0.292	96.1		0.209	96	
5	0.209	98.8		0.208	96.5	
6	0.175	96.8		0.145	97.8	
7	0.139	97		0.169	98.3	
8	0.126	99		0.252	99.1	
9	0.105	99		0.055	98.2	
10	0.109	96.5		0.008	99	

4 Result Discussion

In this paper, a study of weapon detection has been presented. Two different DLMs have been proposed. The first model is based on CNN and the second one is based on a combination of CNN and ConvLSTM. This study is focused on the effect of deploying DLMs on distributed wireless sensor networks for military applications. Hence, the proposed models have been evaluated on the dataset in presence of AWGN and with image compression as the wireless channel affects the images or frames. The proposed scenario is based on image transmission over OFDM as a common modulation technique in wireless communication. Fig. 19 shows a comparison of accuracy between the two proposed models. It can be observed that the proposed hybrid model has a superior performance for weapon detection from images transmitted over a wireless communication channel in the presence of AWGN.

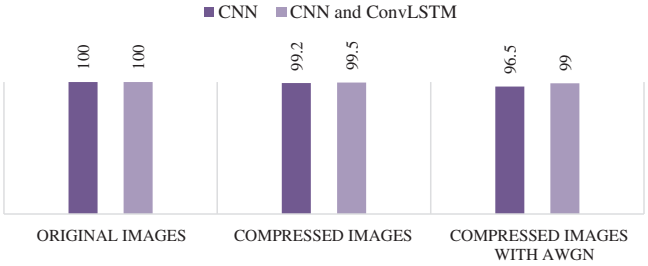

Figure 19: A comparison of accuracy between the two proposed CNN and hybrid CNN and ConvLSTM models

The proposed trend is novel in this field. The contribution of this paper is in the field of detecting carried weapons with soldiers. Furthermore, the proposed DLMs are deployed on the transmitted images from a dynamic environment, which is frequently encountered in military applications. The main trend of research works in the literature is to detect weapons in a static environment. Therefore,

we can consider the proposed models as efficient solutions for military, surveillance and data analysis applications.

5 Conclusions

A significant issue involving weapon detection from IR images has been discussed in this work. The objective is to detect the weapon type from the collected images from drones. Different DLMs have been deployed to achieve this objective. The video stream is compressed using SPHIT encoder, and finally it is broadcasted over OFDM channel to reach the receiver side, where DLMs are applied for object detection. This work has presented a CNN model, and a hybrid model that contains CNV and ConvLSTM layers. The proposed models have been carried out on original, compressed and compressed with AWGN images. The proposed models reveal a superior performance for weapon detection in the presence of robust conditions as they achieved an accuracy of 99%. Furthermore, this work can be improved in the future by adding another stage to segment the objects of interest in the images.

Acknowledgement: The authors would like to thank the support of the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University.

Funding Statement: This research was funded by the Deanship of Scientific Research at Princess Nourah Bint Abdulrahman University through the Research Funding Program (Grant No# FRP-1440-23).

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

References

[1] D. H. Woo, I. K. Nam and H. C. Lee, "Smart reset control for wide-dynamic-range LWIR FPAs," *IEEE Sens. J.*, vol. 11, no. 1, pp. 131–136, 2010.
[2] M. Mahesh, "Airport full-body scanners," *J. Am. Coll. Radiol*, vol. 7, no. 5, pp. 379–381, 2010.
[3] N. Agrawal, S. J. Darak and F. Bader, "New spectrum efficient reconfigurable filtered-OFDM based l-band digital aeronautical communication system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, pp. 1108–1122, 2019.
[4] J. Chen, B. Daneshrad and W. Zhu, "MIMO performance evaluation for airborne wireless communication systems," in *MILCOM 2011 Military Communications Conf.*, Baltimore, Maryland, USA, pp. 1827–1832, 2011.
[5] E. Saberinia and B. T. Morris, "OFDM performance assessment for traffic surveillance in drone small cells," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2869–2878, 2018.
[6] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996, https://doi.org/10.1109/76.499834.
[7] R. Sudhakar, R. Karthiga and S. Jayaraman, "Image compression using coding of wavelet coefficients–a survey," *ICGST-GVIP J.*, vol. 5, no. 6, pp. 25–38, 2005.
[8] H. Zhang, X. Wang, Y. Sun and X. Wang, "A novel method for lossless image compression and encryption based on LWT, SPIHT and cellular automata," *Signal Process. Image Commun.*, vol. 84, pp. 115829, 2020.
[9] K. M. Attia, M. A. El-Hosseini and H. A. Ali, "Dynamic power management techniques in multi-core architectures: A survey study," *Ain Shams Eng. J.*, vol. 8, no. 3, pp. 445–456, 2017.

[10] A. Ben Abdallah, A. Zribi, A. Dziri, F. Tlili and M. Terré, "Adaptive joint source-channel coding using multilevel codes for unequal error protection and hierarchical modulation for SPIHT image transmission," in *IEEE 19th Mediterranean Microwave Symp. (MMS)*, Hammamet, Tunisia, pp. 1–5, 2019.
[11] F. He, "Exploration of multi-node collaborative image acquisition and compression techniques for wireless multimedia sensor networks," *Int. J. Online Biomed. Eng.*, vol. 15, no. 1, pp. 196–208, 2019.
[12] A. Ben Abdallah, A. Zribi, A. Dziri, F. Tlili and M. Terré, "Ultra wide band audio visual PHY IEEE 802.15. 3c for SPIHT-compressed image transmission," in *Int. Symp. on Signal, Image, Video and Communications (ISIVC)*, Cyprus, pp. 59–64, 2016.
[13] G. Raturi, P. Rani, S. Madan and S. Dosanjh, "ADoCW: An automated method for detection of concealed weapon," in *Fifth Int. Conf. on Image Information Processing (ICIIP)*, Shimla, India, pp. 181–186, 2019.
[14] A. Kaur and L. Kaur, "Concealed weapon detection from images using SIFT and SURF," in *Online Int. Conf. on Green Engineering and Technologies (IC-GET)*, Coimbatore, India, pp. 1–8, 2016.
[15] R. Mahajan and D. Padha, "Detection of concealed weapons using image processing techniques: A review," in *First Int. Conf. on Secure Cyber Computing and Communication (ICSCCC)*, Jalandhar, India, pp. 375–378, 2018.
[16] J. Chadha, "Low-cost concealed weapon detection for school environments using acoustic signatures," in *IEEE 16th Int. Conf. on Networking, Sensing and Control (ICNSC)*, Banff, AB, Canada, pp. 358–362, 2019.
[17] R. Hussin, M. R. Juhari, N. W. Kang, R. C. Ismail and A. Kamarudin, "Digital image processing techniques for object detection from complex background image," *Procedia Eng.*, vol. 41, pp. 340–344, 2012.
[18] A. Sedik, A. M. Illyasu, B. A. El-Rahiem, M. E. A. Samea, A. Abdel-Raheem *et al.,* "Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections," *Viruses*, vol. 12, no. 7, pp. 769, 2020.
[19] M. A. Elaskily, H. A. Elnemr, A. Sedik, M. M. Dessouky, G. M. El Banby *et al.,* "A novel deep learning framework for copy-move forgery detection in images," *Multimed. Tools Appl.*, vol. 79, pp. 19167–19192, 2020.
[20] B. A. El-Rahiem, A. Sedik, G. M. El Banby, H. M. Ibrahem, M. Amin *et al.,* "An efficient deep learning model for classification of thermal face images," *Journal of Enterprise Information Management*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JEIM-07-2019-0201.
[21] N. F. Soliman, Y. Albagory, M. A. M. Elbendary, W. Al-Hanafy, E. M. El-Rabaie *et al.,* "Chaotic interleaving for robust image transmission with LDPC coded OFDM," *Wirel. Pers. Commun.*, vol. 79, no. 3, pp. 2141–2154, 2014.
[22] I. Eldokany, E. M. El-Rahaie, S. M. Elhalafawy, M. A. Z. Eldin, M. H. Shahieen *et al.,* "Efficient transmission of encrypted images with OFDM in the presence of carrier frequency offset," *Wirel. Pers. Commun.*, vol. 84, no. 1, pp. 475–521, 2015.
[23] N. F. Soliman, E. S. Hassan, A. H. A. Shaalan, M. M. Fouad, S. E. El-Khamy *et al.,* "Efficient image communication in PAPR distortion cases," *Wirel. Pers. Commun.*, vol. 83, no. 4, pp. 2773–2834, 2015.
[24] F. Jin, P. Fieguth, L. Winger and E. Jernigan, "Adaptive Wiener filtering of noisy images and image sequences," in *Int. Conf. on Image Processing (Cat. No. 03CH37429)*, Barcelona, Spain, vol. 3, pp. III–349, 2003.
[25] N. Singla and N. Singh, "Blood vessel contrast enhancement techniques for retinal images," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 709–712, 2017.
[26] J. Nayak, P. S. Bhat and U. R. Acharya, "Automatic identification of diabetic maculopathy stages using fundus images," *J. Med. Eng. Technol.*, vol. 33, no. 2, pp. 119–129, 2009.
[27] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[28] M. Ranzato, F. J. Huang, Y. -L. Boureau and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 2007, pp. 1–8, 2007.
[29] http://vcipl-okstate.org/pbvs/bench/, last seen at 20 January 2022.