International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

NEXT-GEN AGRICULTURAL SECURITY: INTELLIGENT MULTI-SENSOR SURVEILLANCE WITH DEEP LEARNING-BASED BACKGROUND SUBTRACTION AND CNN CLASSIFICATION.

Himanshu[1*], Harwant Singh Arri[2]
[1*, 2.] Lovely professional university, Punjab, India

ABSTRACT

Large farmers are now begin to apply the new technologies to control their farms. In the new technologies, CCTV cameras are mainly used. These CCTV cameras are used to monitor the farm field manually. At present, some large farmers are using IoT-based systems for full control of the farm field. This system is also capable of identifying stray animals. Such systems are not affordable to any of the small and marginal farmers. This work designs and develops a low-cost intelligent farm field surveillance system to prevent the farm field from stray animals and stealing of farming equipment. A deep learning CNN model is proposed and trained in this work for the classification of objects into animals, suspicious human activities, non-suspicious human activities, and usual motion activities on farms. The proposed CNN model is trained and tested on the specifically developed dataset. The proposed CNN model achieves 92\% average accuracy in the classification of activities with 0.92 precision and 0.91 recall. The proposed CNN model is also compared with well-known deep learning models i.e.MobileNet, VGG16, and ResNet50. The comparison of such models is carried out on the same dataset. The obtained results show that the proposed model outperformed the existing deep learning model in accuracy and precision. The accuracy of the proposed work is 8% higher than the ResNet, 6% higher than the VGG16, and 31% higher than the MobileNet. The obtained results establish the correctness and consistent performance of the proposed model.

Keywords: Detection, Recognition, Background subtraction, CNN

1. Introduction

The majority of the GDP of India depends on the growth of the agriculture sector. However, the growth of this sector is being affected by various factors such as overpopulation, ground data environmental conditions, stray animals, etc. Usually, farmers in India are divided into three different categories, such as marginal, small, and large, based on the size of landholding. Those having less than one hectare, the marginal farmers, constitute a big chunk of the farmers in India. The use of technology with sophisticated farming equipment has been growing with time. The technology is evolving day by day, but in the case of large farmers, access to advanced technology and machines is most among them. They are charging installation and operational costs of such technology so high that it is unaffordable to small and marginal farmers. The crops of small and marginal farmers suffer from stray animals. Such farmers depend on the traditional method of management.
These traditional methods include wire fencing. This wire fencing is not disabled and is unable to stop big animals such as elephants, donkeys, etc. These animals can damage the fencing very easily. Another drawback with this sort of barrier is the high costs, that is, installation costs.
Since plastic fences are available in the market, they also note that the disabled are unable to stop big animals.
Electric fence: These fences are built of metal connected to the electric circuit. At night, the electric current flows in the fence, and this type of fence is capable of avoiding all stray animals, but such a solution is not a humane one. The animal may get high-level injuries. Apart from these, traditional methods such as hatchery waste, smoke, aster oil, etc are also used to protect farm fields from stray animals.

Received: August 02, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Nowadays, big farmers have started using modern technology to manage their farms. In these modern technologies, CCTV cameras are mainly used. These CCTV cameras are used for manual surveillance of the farm field. Some large farmers are using IoT-based systems for the complete management of the farm field. This system is capable of detecting stray animals and crop properties and needs by using soil moisture sensors, humidity sensors, etc. Such systems are too expensive to be afforded by any of the small and marginal farmers.

Alternative to this, an intelligent system with a CCTV camera is needed for such issues. The intelligent system should be capable of the dictation of suspicious and non-suspicious activities in the surveillance zone. Nevertheless, the development of such a system requires a machine learning model to be trained for such a kind of scenario, considering the cost of the system. In this work, a framework for such an intelligent system is proposed to deal with the real-time problems of the farmers in managing their farm fields. This work also proposes a deep-learning-based model for the detection of suspicious activity by humans and intrusion by stray animals. This work proposed an approach for the detection and identification of suspicious and non-suspicious activities, as well as the identification of the presence of animals in the surveillance zone. The proposed work targets suspicious activity such as stealing things or equipment from the farm field, including the crops, etc. The proposed system plans to generate an alarm signal in the farm field to distract the thief or animal, and also generate a notification to the farmer for such incidents. As intelligent capabilities are introduced into the system using deep learning technology, the next chapter discusses the background of deep learning technology

2. Literature review

To recognize any Human activity in the input video, first, there is a need to localize Humans in the given video. Many techniques are there for the detection of Humans and animals. Similarly, for the recognition of Human activity. This talk describes both present approaches.

2.1 Human and Animal Detection

For detection, Figure 1 is a typical process.

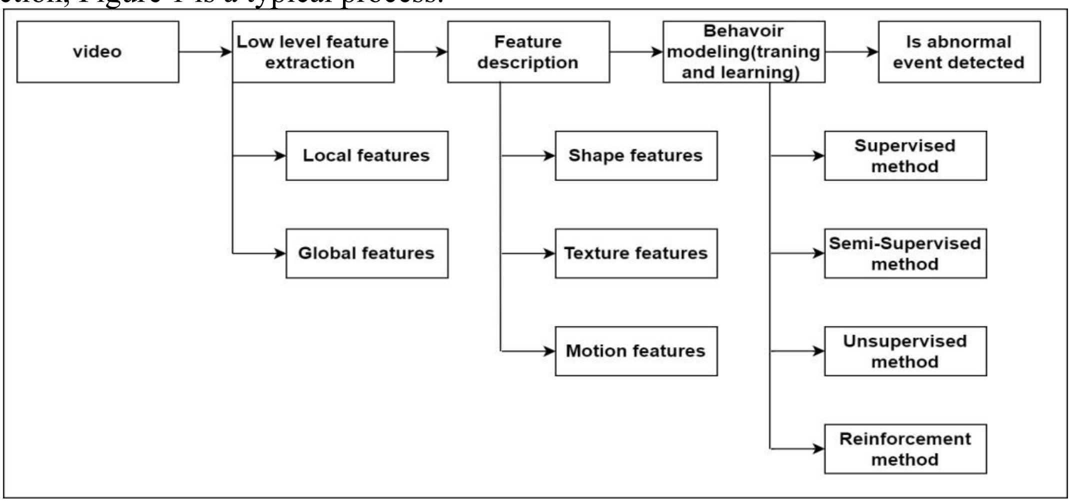


Figure 1. A typical process of event detection

Hussain et al.[1] Proposed a model, which is able to detect moving objects in the video by applying a frame difference approach. Experimental results show that the two frame-difference approaches were better in moving object detection in comparison with a background approach. The added value of this technique is that an object can be located in a complicated background with less amount of background motions very effectively.

Dhulekar et al[2] proposed an approach for human localization and dedication in the frames of any

Received: August 02, 2025

video. The method applies principal component analysis (PCA) to find the potential object regions within the image. Then, a k-nearest neighbors model is applied for the classification of detected object regions as Human or any other object. The method achieves 90-95% accuracy in the proper identification of Human objects from within the videos.

Arroyo et al[3] Proposed an expert system for indoor video surveillance to identify the suspicious behavior of Humans. In this approach, a novel Bob fusion technique purpose and used for the identification of Human objects in the detected objects. After background subtraction, various objects may be detected. In order to reduce the search domain, Human object identification is necessary.

Chen and Harang.[4] Use the optical flow technique for motion detection. In this, to recognize Human motion, the algorithm named adjacency matrix-based clustering is used. This algorithm localizes the Human motion in an image. After that, this detected Human motion was used to recognize Human action.

Buys et al.[5] Proposed to stage an iterative model for the detection of the Human body. In the first stage, the algorithm identifies the region based on the pixel value after the background subtraction. In the next stage, a Kinetic model is proposed to identify the Human region. This kinematic model labels the pixels belonging to the Human body's past. However, the system is too complex for the limited comparative resources.

Sharma and Shah.[6] Proposed a cascade classification-based approach for the detection of animals. The approach is mainly designed for the detection of cows on the highway and roads, which is proposed to be used in vehicles for the prevention of collisions
with the animals. The approach achieves 82-85% accuracy in the detection of animals on the roads.

Nascimento and Marques[7] used blob features with 2D contour analysis of the Blob region for the detection and segmentation of the human object in the image. The method applies to many statistical techniques. Maximally likelihood and support map to compute the 2D shape representation of the detected blob called butter flow.

Antonio et al.[8] Proposed the use of supervised machine learning and extracted features for animal detection. The work developed five different feature extraction windows to detect the animals. These feature extraction windows are based on the colour specification of the pixels. The colour specifications, such as RGB, GRAY, LAB, and ENTROPY, are used to generate the feature extraction windows. Later, k-nearest neighbours and random forest machine learning models trend on the extracted features. The proposed attained a 0.7 F1 score on animal detection.

Forslund and Bjarkefur.[9] Utilize a cascade function for detecting and classifying the vehicles. The methodology proposed a novel set of features for the future, using existing Haas features for identifying the animals.

Pedersen et al.[10] Propose a model for the detection and classification of underwater species. Yolo v2 and Yolo v3 CNNs are exploited in this proposed work. Since they were the best in the tests. Brackis dataset is exploited for training and testing. Gupta et al.[11]Proposed a deep learning-based monitoring system for elephant detection on the railway track, which would reduce the number of elephant-related accidents. The work to be presented is a Convolutional Neural Network model applied to the identification of elephants. Its training and testing are performed on 4200 photos from a public dataset, with an accuracy in the range between 92 and 99 percent.

Santhanam et al.[12]Propose a method to detect animals and alert so that accidents due to animals on the road can be prevented, and in this paper, the author is going to propose a Convolutional neural network model that could be trained and tested on the vast open-source dataset. This model could be further implemented in real time.

Prabhakar et al.[13]Proposed work says that motion detection using frame differencing gives efficient results for object tracking and can be applied to various domains. The study further suggests a method of maintaining a record of item movement, which can be retrieved later to study motion.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.2 Human and Animal recognition

Mabrouk et al.[14] The proposed work explains the behaviour representation and behaviour modelling layers of a video surveillance system are discussed in this study. Firstly, he took a look at the most common feature extraction and description approaches. Then, a deep discussion on several approaches and frameworks of behaviour modelling and categorization was conducted. In addition, the most challenging datasets and assessment metrics for the evaluation of a video surveillance system were provided. Finally, some real-world intelligent video systems were demonstrated.

Wang and Snoussi.[15]Propose a model that can identify the anomalous behaviours in video streams using HOFO (Histogram of optical flow orientation) and SVM (Support Vector Machine) with 97.88% accuracy. Test and train the proposed model on either the PETS or UMN datasets. Both datasets have a lot of types of movies like lawn, indoor, and plaza recorded movies, but have only two classes in these datasets normal events and abnormal occurrences.

Le and Tran.[16]Propose a model to monitor the output and motion template for the recognition of abnormal behavior using the codebook approach in the backdrop modeling and HOG-SVM for the recognition of moving objects and person tracking using a Kalman filter, with an accuracy achievement of 99.5%. In the proposed model, testing and training are performed using the CAVIAR dataset. This data set has recorded some pre-simple videos.

Miao and Song.[17] Propose a hybrid optimization of feature selection and the SVM model, which was implemented in this model. The feature selection optimization algorithm is a combination of the simulated annealing technique and the genetic algorithm, and it implements crossover and mutation probabilities at the same time. This model utilizes parameter tuning for the SVM training model to further improve the detection accuracy of abnormal events caught on security video clips. In experiments, our approach is compared to independent optimization by using GA and PSO separately. The experimental results prove that the hybrid optimization algorithm proposed in this paper can effectively reduce the detection time and improve the classification accuracy for the video surveillance abnormal detection task.

Mu et al.[18] Propose a model that can detect suspicious activity by utilizing the motion vectors in HD videos. The modulus of the motion vectors is further used to deduce the direction and velocity. Lastly, SVM is used to learn and classify the input videos. The results reveal a high recognition rate with a low false alarm rate. In the proposed work KTH(having five different categories)dataset is used for training and testing purposes of the model and achieves 97.5% of accuracy.

Li et al.[19] Model to Detect and Recognize Anomalies of Behavior key to the proposed model is segmented motion patterns. Recently active research fields have arisen around interesting topics given their possible real-world use. The methods were trained and tested on several datasets, including the UCF Crowd Dataset, UMN Crowd Dataset, UCSD Anomaly Detection Dataset, Violent-Flows Dataset, and CUHK Dataset to achieve over 90% accuracy.

Dulhare and Ali[20]suggested an approach that delivers superior results when employing data augmentation techniques to address variations in the underwater environment. Faster R-CNN, coupled with data augmentation. The researchers utilized secondary data for their study, sourcing videos from YouTube. The primary objective is to enhance life-saving measures through early detection of individuals at risk of suffocation underwater.

A model for pneumonia detection using CNN-based feature extraction is proposed by Varshni et al.[21]This model combines SVM and DCNN, a Convolutional Neural Network of 169 layers. The ChestX-ray14 dataset provides the source of secondary data used in this experiment. The same dataset is used for both testing and training. They get an accuracy rate of above 90%. Using a DCNN, Kim and Moon[22] are able to categorize humans for different reasons and also classify human activity, with 97.6% accuracy for human classification and 90.6% accuracy for human activity classification. They use spectrograms and the primary data for that experiment.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Proposed Work

This work proposes an intelligent system for farm field surveillance to prevent the farm field from stray animals and farming equipment from being stolen. Objectives of the proposed work are flowing.

- Development of a deep learning model for detection of animals and suspicious human activities in the farm field.
- A developed deep learning model should be capable of detecting animals from a deep distance as well as closer proximity to the camera point.
- A developed deep learning model should be capable of Identification of activities such as stealing different farming equipment.

The execution of the proposed system requires at least one camera or more depending upon the area of the farm field to be installed at an appropriate location in the farm field. The proposed layout of the farm field surveillance zone set up is shown in figure 2.

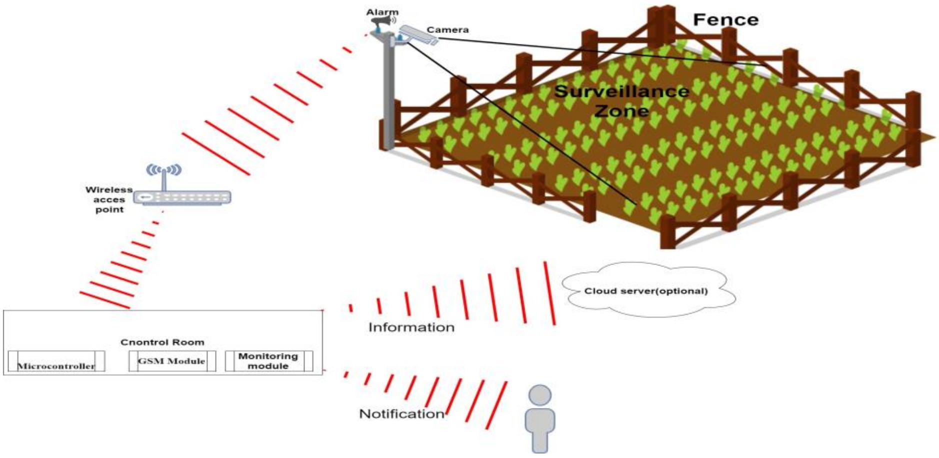


Figure 2. Proposed layout of the farm field surveillance zone setup.

In this, the camera must be connected to the microcontroller either by a wired or wireless medium. The main role of the microcontroller is to process the video stream obtained from the camera(s), identification of any unusual activities, and notify the farmer. An alarm system needs to be equipped with a camera to scare away stray animals and suspicious humans. The microcontroller will be connected to the alarm system. This may also be connected with a cloud data center depending upon the requirement. This microcontroller will raise the alarm in the farm in case of the presence of a stray animal or unusual activity. Parallel to this, the microcontroller will send a notification SMS to the farmer in such a case. An AI-enabled module, a deep learning model, is used on the microcontroller for the detection of stray animals and unusual human activities. The present work proposes a deep learning model for the detection and identification of various activities in the surveillance zone. Figure 3 presents the workflow of the proposed approach.

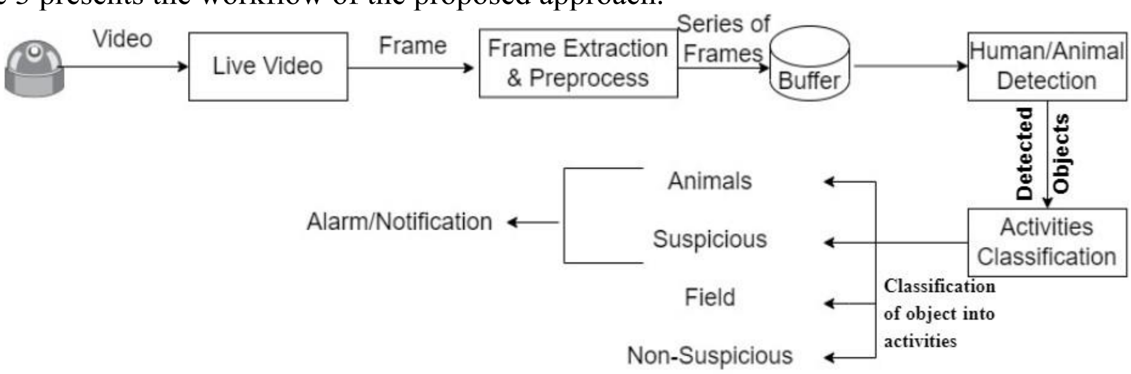


Figure 3. Workflow of the proposed system.

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The proposed approach comprises two main phases, namely Human and animal detection, and activity recognition. In any farm field, there are lots of different activities like the movement of birds, irrigation by a sprinkler system, Human activities, and animal activities. Therefore, the task is performed in two phases. In the first place, during the initial processing phase, the detection of all motion objects is performed. Then, the identification of human and animal objects by using a proposed deep learning model is done. The detail about these two phases is explained in this section.

4. Moving objects detection

First, in this, the video obtained from the camera is processed. Some pre-processing is done, and the frames from the video are stored in the buffer. In this processing, 2-d convolution of the extracted frame is performed by using the Gaussian filter of size 3*3. This operation is done to decrease the noise in the image.

A time-stamp at which the frame is captured is used as a label on the processed frame. Labelling is done to preserve the sequence of the frames in the buffer. This work proposes the use of the background subtraction method for motion detection in the farm field. Thus, the very first frame of the scene is considered as background image of the scene. Also, to be adaptive to the dynamic environment, the background image is changed after every 500 frames.

Let the extracted pre-processed frames stored in the buffer be represented as f_t where t is the time-stamp at which the frame is extracted.

here, the sequence of extracted, pre-processed, and labelled frames is represented as:

$$f_t , f_{t+1}, f_{t+2} \ldots\ldots, f_{t+n} \quad (1)$$

Therefore, the initial background image (b) is defined as:

$$b = f_t \quad (2)$$

In order to detect the moving object in the sequence of frames, a background subtraction operation is applied. Before the background subtraction, both, extracted frame and the background image are converted into grey level images, and these are represented as P_{gt} and $f_{g,}$ respectively. The process of moving object detection is performed as Algorithm 1.

In Algorithm 1,

f_{sub} denotes the resultant image after background subtraction.

f_{bin} represents the binary version of f_{sub} from Otsu's binarization.

f_o is an object vector, i.e. collection of all detected moving objects.

Algorithm 1 Moving Objects Detection and Segmentation

counter ← 1
for <i=2 to condition> **do**
  $fsub = fgt{+}i - bg$
  f_{bin} = thresholding(f_{sub},otsu's threshold)
  f_o = blob detection (f_{bin})
  counter + +
  if (*counter* == 500)*then*
  $bg = fgt{+}i$
  counter == 1
end for

Here,

$$f_o = (O_1, O_2 \ldots\ldots, O_k) \quad (3)$$

Each, O_k has coordinates of the top-left (p1(x1,y1)) and bottom-right(p2(x2,y2)) point of the contour of the object as shown in Figure 4.

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 4. Representation of contour of the object

Therefore, O_k is represented as:

$$O_k = (x_1, y_1, x_2, y_2) \quad (4)$$

After obtaining the object vector P_o, each detected object O_k is isolated from the original frame f_{gt+i} as per the location coordinates defined in O_k and stored in vector f_v. Thus, each detected object is stored as a segmented RGB image and represented as O_{cn}.

$$f_v = (O_{c1}, O_{c2}......., O_{cn}) \quad (5)$$

Figure 5 illustrate the complete process of moving object detection at the first phase.

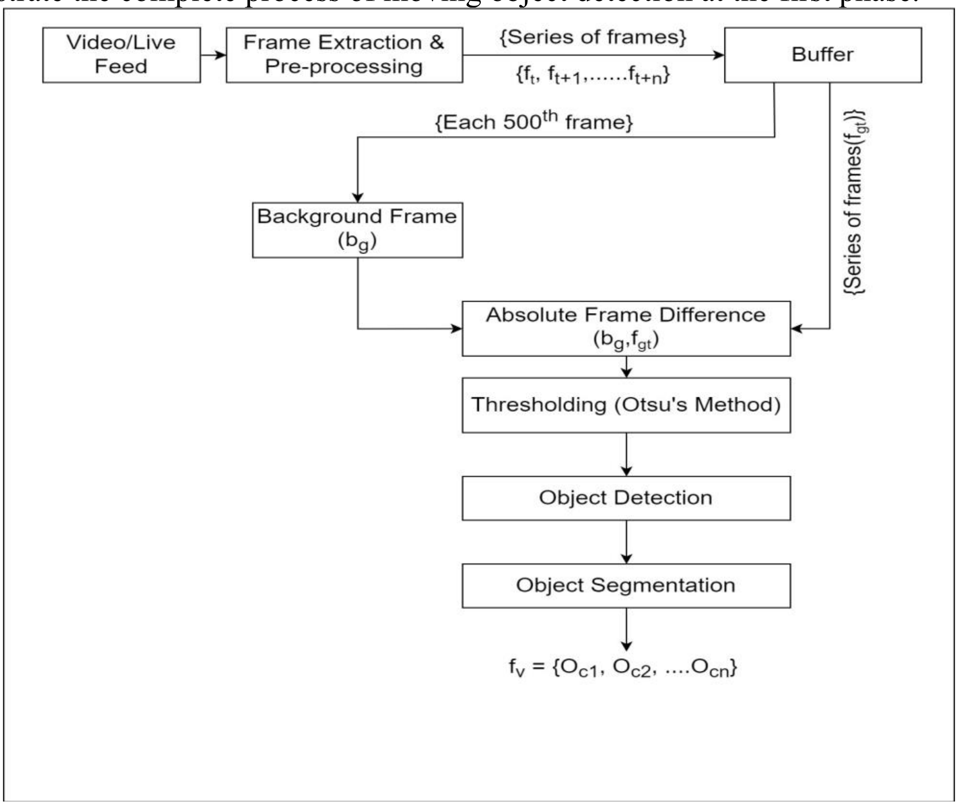


Figure 5. Process of proposed moving objects detection and segmentation

Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. Human Activity and Animal Recognition

In this work, a deep learning model is designed using a Convolutional Neural Network. These deep learning models process each detected object stored in vector f_v in order to classify the activity of the human or animal. Other than predefined categories, all detected objects are discarded, considering the usual activity.

5.1 Proposed CNN Model

The proposed CNN model is composed of a set of eight Convolution+Relu layers, and one pooling layer, along with eight fully connected layers. These layers are shown in Figure 6. This CNN model is designed for an input vector of 64*64*3 size, and the output layer is designed to classify the input among the four classes. The detailed architecture of the proposed CNN model is shown in Figure 6. In this, the size of filters at each convolution layer is kept consistent, i.e., 3x3. The number of filters, such as 64, 128, 256, 128, 64,64,64, and 64 is used from the first convolution layer to the last, respectively. Relu is the activation function used at each convolution layer. A max pooling operation with a 3x3 size window is applied to the activation stack after the initial eight convolution layers. This is an operation that reduces the dimensionality of the activation stack. The features to be extracted will make up the activation stack of this layer. Further, a set of nine fully connected layers is added to this model. In these fully connected layers, from the first fully connected layer to the output layer, 1000, 1000, 780, 780, 800, 800, 500, 500, and 4 perceptron units are respectively used. Relu is the activation function used at each fully connected layer except for the output layer. In the output layer, a Softmax function is used. In order not to over-fit the model, 30% dropouts are applied after the initial fully connected layers.

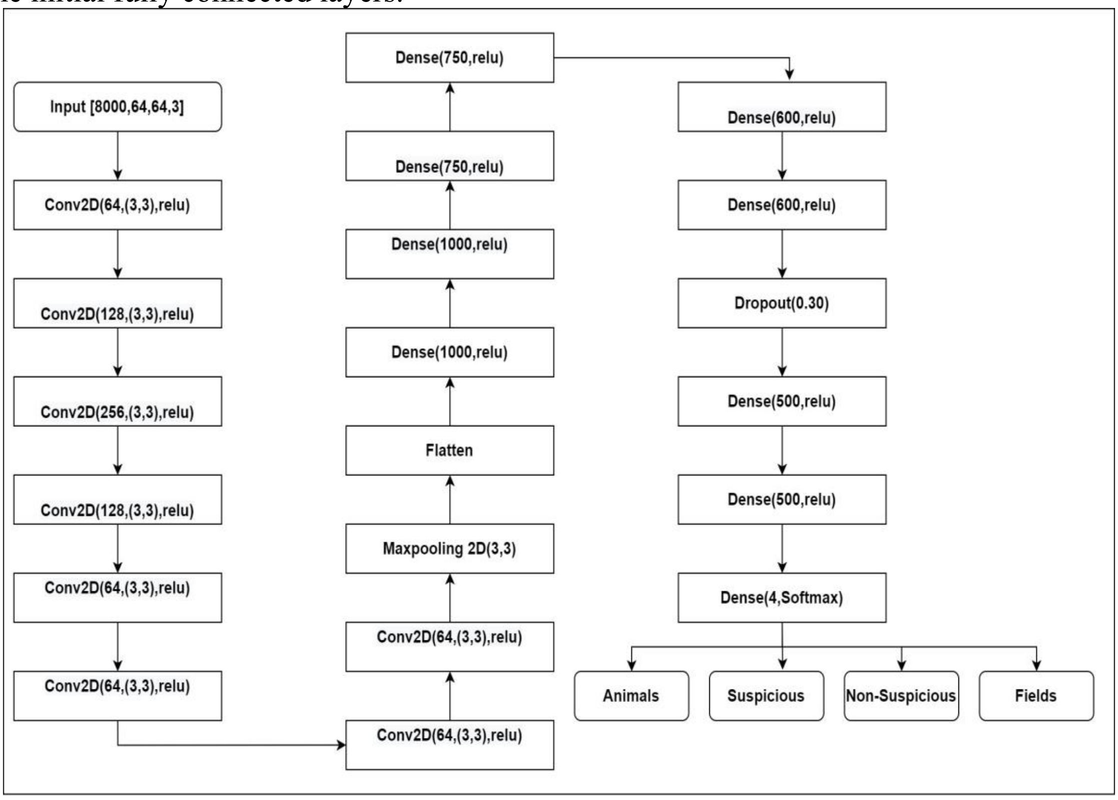


Figure 6. Architecture of the proposed CNN model

The input vector for the CNN is used as shown in Equation 6.

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$X=\begin{bmatrix} X_{1,1}^{1} & - & - & X_{1,N}^{1} \\ - & - & - & - \\ - & - & - & - \\ X_{M,1}^{1} & - & - & X_{M,N}^{1} \\ X_{1,1}^{2} & - & - & X_{1,N}^{2} \\ - & - & - & - \\ - & - & - & - \\ X_{M,1}^{2} & - & - & X_{M,N}^{2} \\ X_{1,1}^{n} & - & - & X_{1,N}^{n} \\ - & - & - & - \\ - & - & - & - \\ X_{M,1}^{n} & - & - & X_{M,N}^{n} \end{bmatrix} \tag{6}$$

Where n is the total number of samples,
N is the number of columns in an input sample,
M is the number of rows in an input sample. To represent the target classes, for digit one-hot encoding is used. That is as follows:

$$y^{T}{}_{i} = (d_1, d_2, \ldots.., d_j) \tag{7}$$

where, for i^{th} class, $d_i = 1$ and $d_j = 0$ $\forall i \neq j$ and $j \in [1,4]$
Thus, the target vector y^T is defined as:

$$y^{T} = \begin{bmatrix} d_1^1 & d_2^1 & - & d_k^1 \\ - & - & - & - \\ - & - & - & - \\ d_1^m & d_2^m & - & d_k^m \end{bmatrix} \tag{8}$$

Here, n is the total number of samples in the input vector X, and k is 4 (Target Classes)
Activation function at each convolution layers and at each fully connected layer relu is defined in equation number 9.

$$\hat{Y} = max(0, y) \tag{9}$$

At the output layer, a softmax is applied to the output of the layer.

$$\hat{Y}_{\mathrm{i}} = \frac{e^{y_i}}{(\sum_{j=i}^{k} e^{y_i})} \tag{10}$$

Where k =4(number of classes in proposed CNN model).
In order to measure the performance of the proposed CNN model, a log loss function is used to evaluate the error between actual and predicted output. The function is shown in equation 11.

$$\mathrm{Loss(E)} = -\sum_{i=1}^{k} y_i^T \log(\hat{Y}_i) \tag{11}$$

Where k is 4(number of classes in the proposed CNN model) and y_i^T is the target output as defined in equation 8.
$\hat{Y}_{\mathrm{I}}$ s the predicted $\mathrm{i_{th}}$ class shown in equation 10.
To minimize the overall loss of the model, the Adam optimization algorithm is used to adjust the parameters of the model. As in many cases, the Adam algorithm is proven to be very fast at minimizing the loss. The cost of function optimization is defined in equation 12.

$$J = -\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{k} y_i^T \log(\hat{Y}_i) \tag{12}$$

Where n is the total number of samples, y^T is the target vector as defined in equation 8, $\hat{Y}_i$ is the

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

predicted i^{th} class shown in equation 10.

A dataset is developed to train the proposed CNN model. The next chapter discusses the training, results, and analysis of the performance of the proposed model.

6. Results and Analysis

In this work, the proposed deep learning model plays a significant role related to the overall performance of the system. This deep learning model is designed to recognize animals and human activities in the farm field. Any deep learning model requires sufficient training on relevant data to achieve good accuracy. A relevant dataset that contains animals, humans in the farm area is required to train the proposed deep learning model. Therefore, a rel-event dataset is developed to train the model.

6.1 Dataset Overview

This dataset has 8000 images of animals, humans with different activities. The overall dataset is divided into four different classes, namely animals, suspicious human activities, non-suspicious human activities, and farm field. In the training set, 75% images of overall dataset are included. Rest 25% images are included in the validation set. Class-wise distribution of images in the training and validation sets is shown in Table 1.

Table 1. Class-wise distribution of images in training and validation sets

Classes	**Number of images in training set**	**Number of images in the validation set**
Animals	1490	510
Field	1516	484
Suspicious	1480	520
Not suspicious	1514	486

All the images in the dataset are recorded in the various farm fields. The proposed CNN model is developed using TensorFlow, Keras Python libraries. The opencv, matplotlib are used for basic image processing tasks. Initially, the training of deep learning model is performed. A separate test of the trained model is also carried out to evaluate the performance of the model.

6.2 Performance analysis of the proposed CNN model in the Training session

In the training session, the training set and the validation set are used. The performance of the CNN model in each iteration is recorded. To measure the performance at the initial level, accuracy and loss are used. In order to reduce the overall cost and adjustments of weights, the primary stochastic gradient descent (SGD) algorithm is used. However, RMSprop and Adam algorithms are also used for some sessions, but the results obtained from SGD are quite good, as shown in Table 2. Therefore, SGD is chosen to train the CNN model. The iteration-wise loss in training and validation of the CNN model is shown in Figure 7.

Table 2. Performance of SGD, RMSProp, and Adam algorithms in Training of CNN Model

Optimizer	**Epochs**	**Loss**	**Accuracy**	**Val loss**	**Val accuracy**
SGD	80	0.7897	0.9394	1.3057	0.7879
RMSProp	70	0.0072	1.0000	0.7212.	0.8350
Adam	22	0.0347	0.9892	0.3171	0.9210

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Loss is defined as the deviation in actual values and the predicted result for the evaluation of the model. It describes how well the specific algorithm models the given data.\

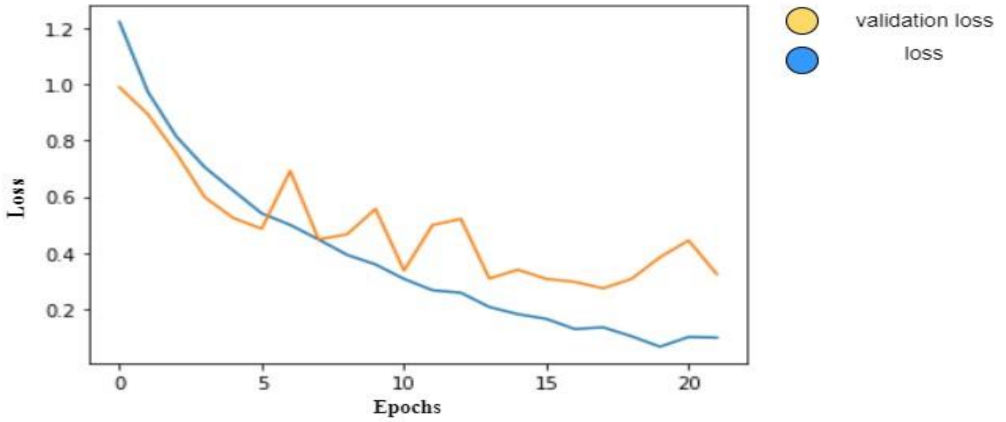


Figure 7. Iteration-wise loss in training and validation

The loss in training is 0.0347 and 0.3171 in validation. Similarly, the accuracy of the model in training and validation in each iteration is shown in Figure 8. Accuracy is the proportion of the total number of correct classifications with respect to the total number of samples. The final accuracy of the proposed model in the training is 98% and in the validation is 92% as shown in Figure 8. The class-wise accuracy of the trained CNN model is shown in Table 3. A comparison of actual and predicted values is also shown in Figure 9.

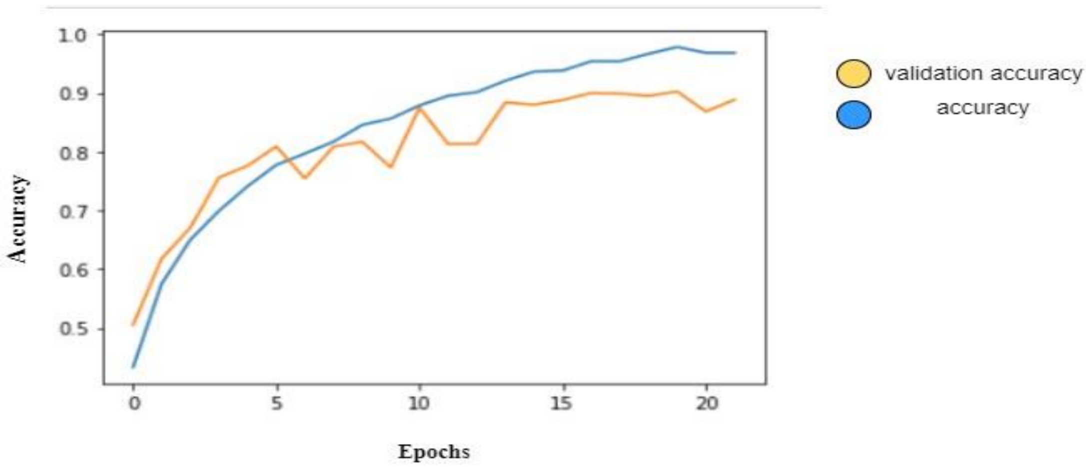


Figure 8. Iteration-wise loss in training and validation

Table 3. Class-wise accuracy of the trained CNN model

Class	Total Number of images	Total Number of Correct Predictions	Accuracy in (%) Accuracy
Animals	510	489	95.8%
Field	484	477	98.5%
Suspicious	520	430	82.6%
Not suspicious	486	446	91.7%

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

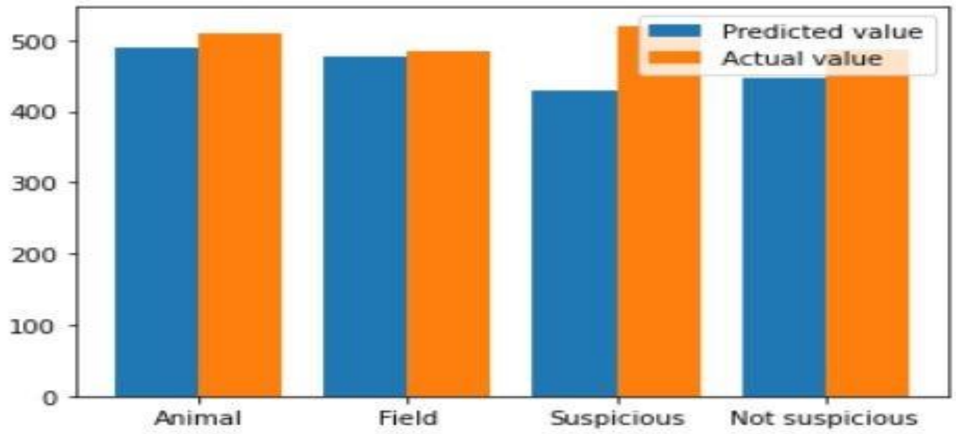


Figure 9. Class-wise Comparison of Predicted and Actual Values

The trained model is able to identify animals in the farm field with 95.8% accuracy. Similarly, the model is able to classify suspicious, non-suspicious, and usual farm field activities with 82.6%, 91.7% and 98.5% respectively.

However, the accuracy in suspicious human activity recognition of the trained CNN model is relatively less. This is because of the minor difference between suspicious and non-suspicious human activities. The same may be visualized by analyzing the confusion matrix shown in Figure 10. This matrix illustrates that the trained model predicts 71 samples from suspicious human activities as non-suspicious activities.

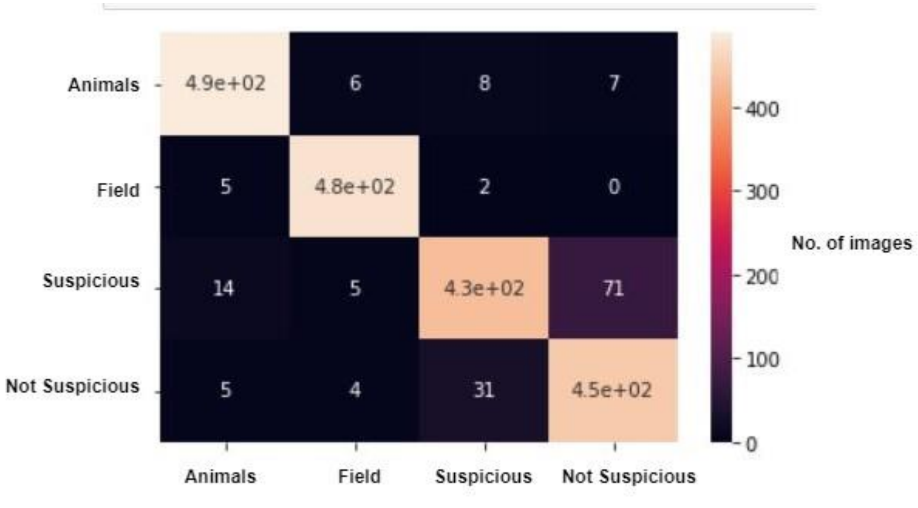


Figure 10. Analysis of results using a confusion matrix

Apart from the accuracy, the performance metrics such as precision, Recall, and f1-score are also evaluated for the trained model. The class-wise results of these metrics are shown in Table 4.

Table 4. Class-wise classification report of the trained CNN

Class	Precision	Recall	f1-score	Support
Animals	0.95	0.96	0.96	510
Field	0.97	0.99	0.98	484
Suspicious	0.91	0.83	0.87	520

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Not suspicious	0.85	0.92	0.88	446
Average	0.92	0.92	0.92	

The trained CNN model achieves 0.92 average precision value in the classification. This establishes the consistency of the proposed model in correct classification. The average recall value of the trained CNN model is 0.92, which proves the correctness of the model. Hence, the F1-score of the trained model is 0.92, which is satisfactory.

6.3 Evaluation of the Performance of the Trained CNN Model in Testing

A separate testing session is carried out to evaluate the performance of the trained model on unseen data i.e, data not used in training and validation. Therefore, a second dataset is also prepared for testing purposes. This dataset contains 3993 images from all four categories. The class-wise distribution of images in this dataset is shown in Table 5.

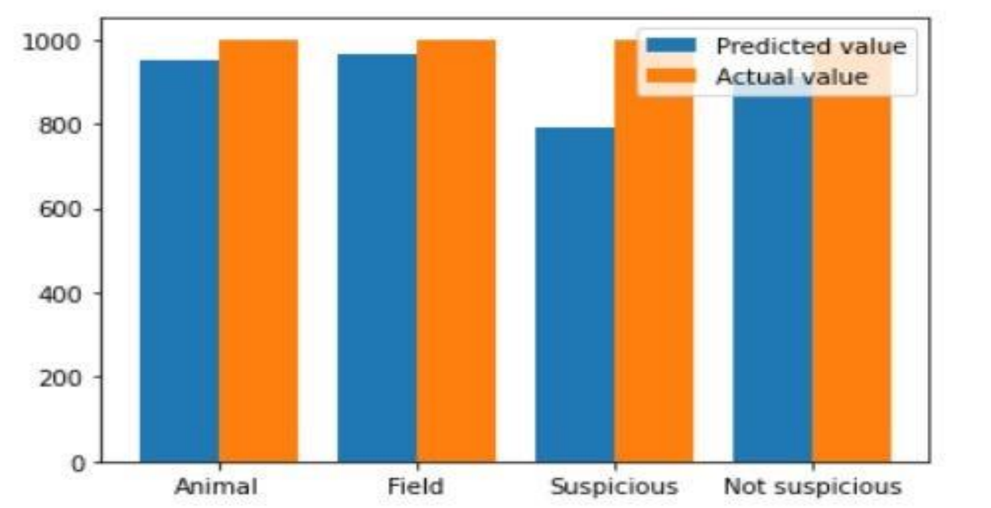
Table 5. Class-wise distribution of images in the testing dataset

Class	**Number of images**
Animals	1000
Field	1000
Suspicious	1000
Not suspicious	993

Classification results of the trained CNN model on the testing dataset are shown in Table 6. The model achieves 90.6% average accuracy on the testing data. The accuracy of the proposed model in animal classification is 95% whereas the accuracy in suspicious human activities recognition it achieves 79.2%. However, it is relatively low compared to the accuracy in other categories. Comparison of actual and predicted values for each class is shown in Figure 11.

Table 6. Class-wise prediction results of the trained CNN in testing

Class	**Predicted value**	**Actual value**	**Accuracy**
Animals	950	1000	95%
Field	967	1000	96.7%
Suspicious	792	1000	79.2%
Not suspicious	910	993	91.6%



Received: August 02, 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 11. Class-wise Comparison of Predicted and Actual Values in Testing

The precision, recall, and f-score of the model on testing data are shown in Table 7. The model achieves satisfactory results on testing data, as well. The average precision value achieved by the model is 0.91. The model achieved an average recall value, i.e., 0.90. Therefore, the overall results establish the correctness and consistency of the proposed mode.

Table 7. Classification report of the trained CNN in testing

Class	Precision	Recall	f1-score	Support
Animals	0.94	0.95	0.94	1000
Field	0.96	0.97	0.96	1000
Suspicious	0.90	0.79	0.84	1000
Not suspicious	0.83	0.92	0.87	993
Average	0.91	0.90	0.90	

6.4 Comparison of the proposed CNN model with existing deep learning models

The proposed work is compared with well-known deep learning models such as MobileNet[23], VGG16[24], and ResNet50[25]. In this work, all models are customized to classify the input into the given four categories. The performance of these deep learning models is tested on the testing dataset proposed in this work. The obtained results are shown in Table 8. In this experiment, the proposed work achieves relatively good accuracy and precision values as compared to existing models. The accuracy of the proposed deep learning model is 8%,6%, and 31% higher than the ResNet50, VGG16, and MobileNet, respectively. The precision value of the proposed work is 2%, 6%, and 32% higher than the ResNet50, VGG16, and MobileNet respectively. The correctness of the suggested model is contrasted with several other authors' models in Table 9.

Table 8. Comparison with different deep learning models

Model name	Accuracy	Precision	Recall	f1-score
ResNet50[25]	83%	89%	91%	91%
VGG16[24]	85%	85%	85%	85%
MobileNet[23]	60%	59%	59%	59%
Proposed model	91%	91%	91%	91%

Table 9. Comparison with different deep learning models

Name	Methodology used	Accuracy
Foreground Detection[26]	Gaussian Mixture Models (GMM)	85%
Deep Learning-Based Detection[27]	MTCNN	90%
Multimodal Deep Learning[28]	MDLF	86.26%
Proposed model	Proposed CNN+Background subtraction	91%

7. Conclusions and Future Scope

This work proposes a low-cost intelligent farm field surveillance system to prevent the farm field from stray animals and the theft of farming equipment. The proposed approach is equipped with an object detection module, an animal detection and human activities recognition module, and an action module. The object detection module detects the moving objects in the farm field by processing the live video stream obtained from the camera. The next module identifies animals and human activity from detected objects. The action models start alarming and notification of the farmer in case of the detection of any stray animals or suspicious human activity. A deep learning CNN model is proposed

and trained in this work for the classification of objects into animals, suspicious human activities, non-suspicious human activities, and usual motion activities in farm. The proposed CNN model is trained and tested on the specifically developed dataset. The proposed CNN model achieves 92% average accuracy in the classification of activities with 0.92 precision and 0.91 recall. The proposed CNN model is also compared with well-known deep learning models i.e.MobileNet, VGG16, and ResNet50. The comparison of such models is carried out on the same dataset. The obtained results show that the proposed model outperformed the existing deep learning model in accuracy and precision. The accuracy of the proposed work is 8% higher than the ResNet, 6% higher than the VGG16, and 31% higher than the MobileNet.
The obtained results establish the correctness and consistent performance of the proposed model.
The approach proposed in this work achieves overall good results. However, the accuracy of the proposed work is relatively less in the classification of suspicious human activities. As the model predicts some of the samples into the non-suspicious category. In the future, the proposed deep learning model can be fine-tuned. The model can be further extended to classify specific human activities.

References

1. Hussain, Z., Naaz, A. & Uddin, N. Moving object detection based on background subtraction and frame differencing technique. *Int. J. Adv. Res. Comput. Commun. Eng.* 5, 817–819 (2016).
2. Dhulekar, P., Gandhe, S., Chitte, H. & Pardeshi, K. Human action recognition: an overview. In *Proceedings of the international conference on data engineering and communication technology*, 481–488 (Springer, 2017).
3. Arroyo, R., Yebes, J. J., Bergasa, L. M., Daza, I. G. & Almazán, J. Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert. systems with Appl.* 42, 7991–8005 (2015).
4. Chen, D.-Y. & Huang, P.-C. Motion-based unusual event detection in human crowds. *J. Vis. Commun. Image Represent.* 22, 178–186 (2011).
5. Buys, K. *et al.* An adaptable system for rgb-d based human body detection and pose estimation. *J. visual communication image representation* 25, 39–52 (2014).
6. Sharma, S. U. & Shah, D. J. A practical animal detection and collision avoidance system using computer vision technique. *IEEE Access* 5, 347–358 (2016).
7. Nascimento, J. C. & Marques, J. S. Performance evaluation of object detection algorithms for video surveillance. *IEEE Transactions on Multimed.* 8, 761–774 (2006).
8. Antônio, W. H., Da Silva, M., Miani, R. S. & Souza, J. R. A proposal of an animal detection system using machine learning. *Appl. Artif. Intell.* 33, 1093–1106 (2019).
9. Forslund, D. & Bjärkefur, J. Night vision animal detection. In *2014 IEEE intelligent vehicles symposium proceedings*, 737–742 (IEEE, 2014).
10. Pedersen, M., Bruslund Haurum, J., Gade, R. & Moeslund, T. B. Detection of marine animals in a new underwater dataset with varying visibility. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 18–26 (2019).
11. Gupta, S., Mohan, N., Nayak, P., Nagaraju, K. C. & Karanam, M. Deep vision-based surveillance system to prevent train–elephant collisions. *Soft Comput.* 26, 4005–4018 (2022).
12. Santhanam, S., Panigrahi, S. S., Kashyap, S. K., Duriseti, B. K. *et al.* Animal detection for road safety using deep learning. In *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 1–5 (IEEE, 2021).
13. Prabhakar, N., Vaithiyanathan, V., Sharma, A. P., Singh, A. & Singhal, P. Object tracking using frame differencing and template matching. *Res. J. Appl. Sci. Eng. Technol.* 4, 5497–5501 (2012).
14. Mabrouk, A. B. & Zagrouba, E. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert. Syst. with Appl.* 91, 480–491 (2018).

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

15. Wang, T. & Snoussi, H. Detection of abnormal visual events via global optical flow orientation histogram. *IEEE Transactions on Inf. Forensics Secur.* 9, 988–998 (2014).
16. Le, T.-L. & Tran, T.-H. Real-time abnormal events detection combining motion templates and object localization. In *The National Foundation for Science and Technology Development (NAFOSTED) Conference on Information and Computer Science*, 17–30 (Springer, 2014).
17. Miao, Y. & Song, J. Abnormal event detection based on svm in video surveillance. In *2014 IEEE workshop on advanced research and technology in industry applications (WARTIA)*, 1379–1383 (IEEE, 2014).
18. Miao, Y. & Song, J. Abnormal event detection based on svm in video surveillance. In *2014 IEEE workshop on advanced research and technology in industry applications (WARTIA)*, 1379–1383 (IEEE, 2014).
19. Li, T. *et al.* Crowded scene analysis: A survey. *IEEE transactions on circuits systems for video technology* 25, 367–386 (2014).
20. Dulhare, U. N. & Ali, M. H. Underwater human detection using faster r-cnn with data augmentation. *Mater. Today: Proc.* 80, 1940–1945 (2023).
21. Kim, Y. & Moon, T. Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks. *IEEE geoscience remote sensing letters* 13, 8–12 (2015).
22. Khan, S. H. *et al.* Covid-19 detection and analysis from lung ct images using novel channel boosted cnns. *Expert. Syst. with Appl.* 229, 120477 (2023).
23. Pujara, A. Image classification with mobilenet. https://medium.com/analytics-vidhya/ image-classification-with-mobilenet-cc6fbb2cd470/ (2020). Published in jan 04, 2020.
24. Rohini. Everything you need to know about vgg16. https://medium.com/ @mygreatlearning/ everything-you-need-to-know-about-vgg16-7315defb5918/ (2021). Published in sep 23, 2021.
25. Dwivedi, P. Understanding and coding a resnet in keras. https://towardsdatascience.com/ understanding-and-coding-a-resnet-in-keras-446d7ff84d33/ (2019). Published in jan 04, 2019.
26. Amin, A., Anzum, M. F. & Mondol, M. H. *Abnormal behavior detection of human by video surveillance system*. Ph.D. thesis, BRAC University (2014).
27. Irshad, T. *et al.* A deep learning based human detection and tracking for security surveillance systems. *Appl. Comput. Eng.* (2023).
28. Yuvasini, D., Jegadeesan, S., Selvarajan, S. & Mon, F. A. Enhancing societal security: a multimodal deep learning approach for a public person identification and tracking system. *Sci. Reports* 14, 23952 (2024).

Author contributions statement

All authors have participated equally in conception and design, or analysis and interpretation of the data, drafting the article and approval of the final version.

Data Availability

Due to privacy and ethical concerns, neither the data nor the source of the data can be made publicly available. However it can be available on request from the corresponding author.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.