Machine Learning-Based Intelligent Video Surveillance in Smart City Framework

Mohammed A. J. Maktoof*[1], Ibraheem H. M.[2], Mohammed A. Abdul Razzaq[3], Ahmed Abbas[4], Ali Majdi[5]

[1] Al-Turath University College, Baghdad, 10021, Iraq
[2]Department of Computer Techniques Engineering, Al-Rafidain University College, Baghdad 10064, Iraq
[3]Department of Computer Techniques Engineering, Mazaya University College, Thi Qar, Iraq
[4]Department of Medical instruments engineering techniques, Alfarahidi University, Baghdad, Iraq
[5]Department of Buildings and Construction Techniques Engineering, Al-Mustaqbal University College, 51001 Hillah, Babylon, Iraq
Emails: mohammed.maktof@turath.edu.iq; ibraheem.hatem.elc@ruc.edu.iq; mohammed.wahab@mpu.edu.iq; ahmed.abbas@alfarahidiuc.edu.iq; alimajdi@uomus.edu.iq

Abstract

The proposed method of using Machine Learning in Motion Detection and Pedestrian Tracking-assisted Intelligent Video Surveillance Systems (ML-IVSS) can be seen as an application of intelligent fusion techniques. ML-IVSS combines the power of motion detection, pedestrian tracking, and machine learning to create a more accurate and efficient surveillance system for smart cities. By fusing these techniques, ML-IVSS can effectively detect unusual behaviors such as trespassing, interruption, crime, or fall-down, and provide accurate depth data from surveillance footage to protect residents. Intelligent fusion techniques can help improve the accuracy and effectiveness of surveillance systems in smart cities, making them safer and more secure for residents. Combination channel models are used at first, and an object area with prominent features is selected for surveillance. Scaled modification and extraction of features are carried out on the presumed object's region. Identifying the low-level characteristic is the first step in incorporating it into neural architectures for deep feature learning. A smart CCTV data set is used to evaluate the proposed method's performance. According to the numerical analysis, the proposed ML-IVSS model outperforms other traditional approaches in terms of abnormal behaviour detection (98.8%), prediction (97.4%), accuracy (96.9%), F1-score (97.1%), precision (95.6%), and recall (96.2%).

Keywords: Video Surveillance System; Machine Learning; Smart City; Intelligent Fusion Techniques; Deep Feature Learning.

1. Introduction

Smart cities use various innovative technologies to improve people's quality of life. In recent years, human activity recognition from a video has been undertaken with machine learning and computer vision methods [1]. Violence recognition is a task that can be used in real-life applications. The major tasks of large-scale surveillance models utilized in institutions like schools, prisons, and psychiatric care services make alarms of possibly dangerous circumstances [2]. Yet, security guards are often burdened with many cameras, where manual response periods are regularly high, resulting in a robust demand for automatic alert models [3]. This system must be very effective because many surveillance cameras must be handled. Likewise, there is a growing demand for automatic ratings and tagging systems to process huge videos uploaded to the website [4]. Since smartphones are

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

frequently utilized for record-beating, effective smartphone employment is anticipated too. Intelligent video surveillance offers cutting-edge smart home security that records criminal activity in homes, businesses, and more based on the user's preferences [5].

Anomalous event recognition for video series is a complex challenge due to the instability of the descriptions of abnormality and normality and the dependence of the classifications on the context setting [6]. Presently, the extraction of features is observed as one main factor for anomalous event recognition in existing methods [7]. From the feature depiction point of view, the strange event recognition model is primarily considered the hand-crafted feature-based model and deep feature-based methods [8]. Artificial intelligence aids detect anomalies like a person entering a restricted zone or abnormal behavior and reporting it to the model [9]. In video surveillance, video analytics utilized Deep Learning and Machine Learning approaches to determine objects, classify them, and identify their properties [10].

One of the most important services in smart cities is surveillance video service (SVS). Valued elements, such as trajectories or the visual presence of moving objects, are extracted from large surveillance footage and serve as the foundation for smart traffic monitoring and public security [11]. Explaining latent data effectively through sophisticated video analysis tools must be carefully designed [12]. [25][26][27] Utilizing effective machine learning techniques to analyze the diverse data tangled up in the surveillance video is one of the most important strategies to implement [13]. There are a lot of frames (imageries) in surveillance videos [14]. Any video analysis procedure starts with the raw frames. In videos, it's typical for succeeding frames to differ only slightly from one another and contain repetitive information. As a result, video analysis on these raw frames takes time [15].

Through advances in Computer Vision and Artificial Intelligence (AI) came the use of Edge Computing and Embedded elements, which enhanced the incorporation of Intelligent Surveillance [16]. This development formed a robust relationship between the internet of things and edge computing. Artificial intelligence systems are becoming improved, leading to many improvements in the applications of modern technologies [17]. In video surveillance, existing artificial intelligence and deep learning algorithms are primarily utilized for video analysis [18]. Most current video surveillance systems depend on conventional integration, fronting enormous data announcements above, severe packet loss, and high latency confines [19].

As computer vision progresses, it is now feasible to collect various picture data types using a variety of vision sensors, like a fixed camera or a stereo camera. Since the acquired scene may be aware of context, like item spatiotemporal condition changes, it makes sense (e.g., human behaviors). For example, the new development of competitive RGB sensors has produced an innovation in computer vision's deepest problems, such as the removal of background and disruption from the light sources or other obstructions [20].

The contribution of the work as

- To Design the ML-IVSS model for smart city surveillance and citizens' safety.
- Determining high-level semantic data from raw data and extracting the suitable features for abnormal behavior detection in smart city
- The simulation study shows that the suggested ML-IVSS model improves F1-score ratios, accuracy, and prediction.

The repose of the work is prearranged as trials: Section 2 discourses the literature survey. In section 3, the ML-IVSS model is proposed. In section 4, the proposed model analysis is implemented. Finally, section 5 determines the research article.

2. Literature Survey

Sanjeev Kumar Angadi et al. [21] proposed the Hybrid Bayesian Approach (HBA) for Video Surveillance System. This proposed HBA has two stages. First, the object features were used to identify people, and the spatial features were used. The Viola-Jones method is used to identify objects in the videos at the start of the process. Once the objects are located, a hierarchical skeleton is used in the feature extraction procedure to extract the desired characteristics efficiently. Because it's an effective and understandable abstraction, object skeletons are useful for object identification and matching. The object-based characteristics of the Bayesian network are modified to identify people using the Bayesian network. Only spatial characteristics are retrieved and used in the gait-based Bayesian network to identify people at the stage of spatial-based person identification.

The area under the receiver operating characteristic curve maximization (AUCM) was recommended by Asghar Feizi [22] for video surveillance by modeling normal behaviors. In the

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

early training phases to recognize aberrant behavior, finding a model that reliably predicts typical actions is a significant issue. The histogram-oriented flow (HOF) is retrieved as local-based characteristics as the initial stage in estimating this model. The indicated characteristics of speed, trajectory variance, and deviation from the trajectory are retrieved object-based characteristics. Hierarchical abnormal-behavior identification is proposed here to select behavioral traits in the study's testing phase.

For moving vehicle detection (MVD) systems, Ahilan Appathurai et al. [23] discussed artificial neural networks (ANN) and oppositional gravitational search optimization methods (ANN-OGSA). The two main phases of the proposed system are background generation and vehicle detection. They develop a quick method to produce a neutral background first. Once the backdrop has been constructed, a moving vehicle is detected using the ANN-OGSA model. They use the OGSA method to determine the optimum weight value for the ANN classifier to improve its performance. They demonstrated the system's efficacy using various algorithms and three distinct kinds of videos for analysis.

T. Thenmozhi and A.M. Kalpana carefully considered background subtraction and improved sequential outline separation strategies (BS-ISOSS) [24] to distinguish video datasets. This study recognizes moving things in complex settings with varying item size and quantities. For the simulation, MATLAB is used to calculate the suggested approach's detection error and processing time. Background removal and foreground detection are used for motion and object identification in the sequential outline separation technique presented. This process separates the area of excitement from the developing background. Their suggested methods detect moving objects productively, as shown by simulation results and error rate investigations. Their suggested adaptive motion estimation and sequential outline separation approach outperformed traditional accuracy, sensitivity, and specificity systems.

In order to address the shortcomings of the current approach for efficient smart city surveillance, this research suggests the ML-IVSS model. The proposed model is briefly discussed in the section that follows.

Machine Learning assisted Intelligent Video Surveillance Systems (ML-IVSS)

Smart cities use CCTV surveillance systems to offer public services and protect residents, such as crime prevention and response, forensic evidence, etc. CCTV is a system of cameras dispersed and linked to a central hub, and the images from these cameras are then shown on monitors. Smart cities utilize surveillance technologies to enhance urban mobility and traffic management, making the streets safer and more effective for everyone. Furthermore, by combining real-time and historical traffic data, authorities may anticipate peak periods and make necessary preparations in the future. In particular, a smart city's surveillance monitoring systems have many essential characteristics. These safety features include face recognition, motion detection, and activity tracking, which help reduce the likelihood of a catastrophe. However, automated analysis in video surveillance is a field that may be further studied, particularly in terms of the connection of security equipment. In the security industry, conventional manual monitoring of CCTV (closed-circuit television) images continues to be used. Identifying abnormal behavior in videos is a hotspot for smart security research. Abnormal actions are defined differently in various surveillance video scenarios and surveillance items. Identifying unusual activity among a crowd in public settings has significant research value.

Deep CNN has consistently shown inspiring performance in many vision tasks, particularly object recognition and detection, in the last decade. Conventional movement recognition models mostly use pedestrian detection and segmentation, extracting gesture and motion features for recognition and classification. The success of such models appears to be ascribable to their ability to extract higher-level features via their multiple-layer structure. The perception of high-energy activities is reflected in kinematic patterns of movement and dynamic characteristics. The most important information for distinguishing between conflict and violence is conveyed via motion. Movement information is maybe even more critical in this task than appearance.

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

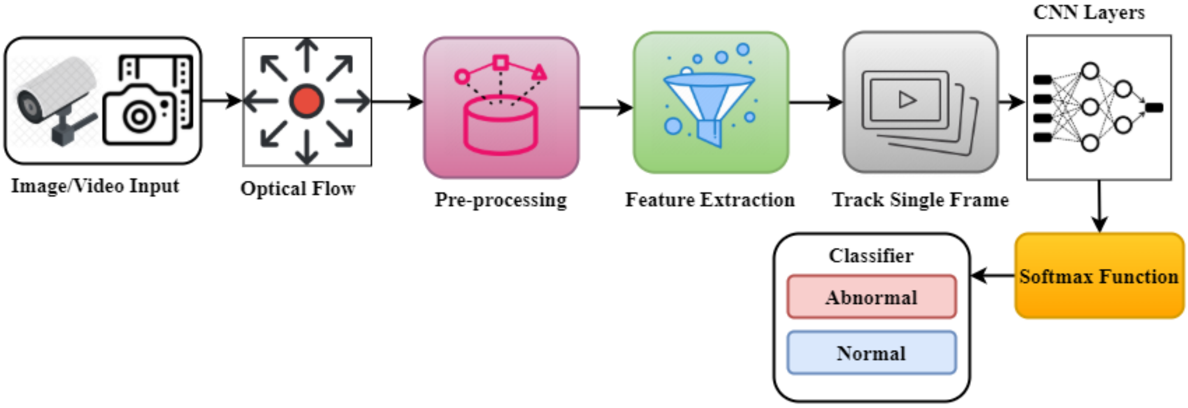

Figure 1: Proposed ML-IVSS model

Figure 1 shows the proposed ML-IVSS model. Individuals pay more and more attention to abnormality recognition, and there are more and more approaches. Generally, the direction and speed of the crowd area like. However, individuals will run away speedily because of fear of evading possible danger when an abnormal event happens. Yet, the abnormal behavior of the crowds has the features of fast motion speed, an unexpected rise in accelerations, and clear concentration of motion in the same direction or balances in manifold directions, high motion range, panic expression, great paces, and chaotic trajectories.

Calculating characteristics like acceleration, speed, direction, and movement amplitude are comparatively simple and expressed by optical flows. The feature extraction, like expressions and steps, is more complex. This article extracts the features of the moving target's acceleration, speed, direction, trajectory, and motion amplitudes to identify anomalous crowd behavior. Conventional anomalous behavior detection models only utilize RGB imageries as the network's input without the hidden movement data in the video series. This article suggests a video anomalous behavior recognition model reliant on a dual-stream CNN to overcome this issue.

A foreground movement block can efficiently signify the motion data of pedestrians. Set the ith space block to A_i, $1 \leq i \leq S \times G$. The foreground movement block denotes the space blocks where the premotion scenic spots perform. The foreground data controlled in a spatial block in the foreground movement blocks can be data like noise, which cannot properly portray the movement object behavior. This article pre-processes every spatial block to identify the presence of such movement blocks. Supposinga_iDenotes the number of front spots in blocks; only when expressed (1) is fulfilled could it be reserved as foreground movement blocks:

$$C_j = A_i, \ if \frac{a_i}{K \times K} > \lambda \qquad (1)$$

The above operation is pre-processing foreground images, and the foreground movement block can be extracted from the pre-processed foreground images. $C_j(1 \leq j \leq W)$ denotes that image blocksA_i It can be utilized as the ith foreground movement blocks when the above states are encountered. $(0.1 \leq \lambda \leq 0.3)$ are the comparison threshold of the prior scenic spots.

The optical flow vectors of every pixel in pre-processed spatial blocks are extracted and mean values are utilized as the optical flow vector of present blocks. Equation 2 utilized as the moving depiction of foreground movement blocks:

$$c_j = \frac{1}{I} \sum_i h_j^i \quad (2)$$

As inferred from equation (2) where $c_j$signifies optical flow vectors of the jth foreground movement blocks. Idenotes the number of every pixel in foreground blocks. h_j^i denotes the optical flow vectors of the ith pixel in the jth foreground movement blocks. $||c_j||$ and Γc_j Symbolize the direction and magnitude, correspondingly, of optical flows of the jthforeground movement blocks.

In line with the determined sub-block effects dimension, the sub-block effects of normal and abnormal behaviors can be efficiently differentiated. Thus, this article utilizes the foreground movement effects of the map feature to illustrate the movement effect of adjacent foreground blocks on space-time blocks.

For spatial blocksA_i and a foreground movement block C_j, to estimate whether the foreground movement blocks affect the nearby space blocks, two index parameters are described:

$$\phi_{ji}^d = \begin{cases} 1, & if \ dist(j,i) < \gamma_d \\ 0, & else \end{cases}$$

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

$$\phi_{ji}^{\theta} = \begin{cases} 1, & if -\frac{\pi}{2} < \theta_{ji} < \frac{\pi}{2} \\ 0, else \end{cases} \quad (3)$$

As shown in equation (3), where $dist(j,i)$ signifies the Euclidean distance among the foreground movement block C_j and space blocksA_i. γ_d Indicates distance threshold. θ_{ji}denotes the angle among the vector from foreground blocksC_j to space blocksA_i and the optical flows of C_j. $(-(\pi/2),(\pi/2))$ denotes the view field when the foreground movement blocks are in movement. These two indexes parameters extend whether space blocksA_i is in the impact range of foreground blocksC_j.

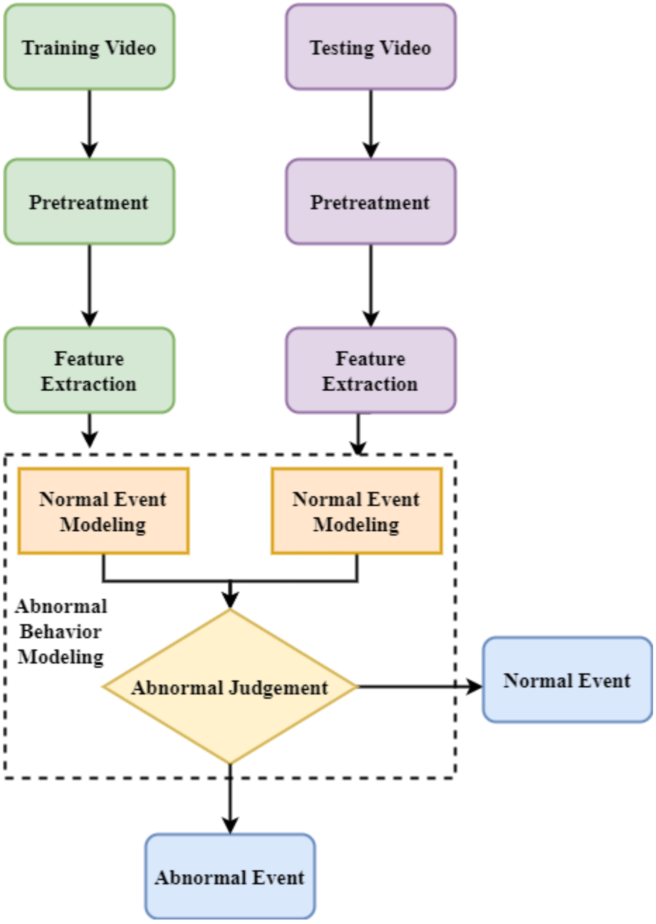

Figure 2: Process Flow of Abnormal Behavior Detection in Smart City

Figure 2 demonstrates the Process Flow of Anomalous Behavior Detection in a Smart City. The video data are pre-processed initially. In this phase, a Gaussian background technique is primarily utilized to remove the background of every frame of a provided picture. In the additional phase, the foreground image characteristics of every space-time block picture are extracted. In the third phase, softmax clusters the feature-extracted datasets. In the fourth phase, an outlier is determined as consistent with the clustering outcomes to identify abnormal image frames. The fundamental-depiction technique of utilizing the space-time blocks as object recognition is not completed to extract movement data from the present space-time blocks directly and rather to analyze further the movement effect of space-time blocks with high foreground data on nearby space-time blocks.

The effect weight of the foreground movement blocksC_j on spatial blocksA_i is described as

$$\omega_i = \begin{cases} \phi_{ji}^{d}\phi_{ji}^{\theta} \exp\left(-\frac{dist(j,i)}{\|c_j\|}\right), & if\, C_j \neq A_i \\ \phi_{ji}^{d}\phi_{ji}^{\theta} \exp\left(-\frac{1}{\|c_j\|}\right), else \end{cases} \quad (4)$$

When space blocksA_i is in the impact range of foreground blocksC_j, the weight of the effect of A_i received by C_j is contrariwise relational to the distance between the two and directly relational to the optical flows magnitude of C_j . $C_j$signifies a pedestrian. When the pedestrian runs forcefully, the weight ω_{ji}rises with the optical flow magnitude. Manifold foreground blocks with various

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

movement directions have contrarily weighted weight on A_i, so all effect weight of $A_i$should be calculated to form an efficient feature depiction. To enhance the degree of the judgment of the extracted feature to compute the effectiveness of the model, the affecting direction Γc_j of foreground movement blocksC_j is quantized by:

$$p(\Gamma c_j) = l_j,\ \ if\ (l_j - 1) \times \frac{2\pi}{q} < \Gamma c_j \leq l_j \times \frac{2\pi}{q} \quad (5)$$

As derived in equation (5) where $l_j \in \{1,2, \dots q\}$ and q is the overall number of quantization direction intervals. l_j Signifies the quantization way index values of the optical flows of the jth foreground movement blocks. The histogram statistic of the effect weight produced by the foreground blocks for spatial blocks A_i defined as:

$$f_i(l_j) = \sum \omega_{ji}\ ,\ if\ p(\Gamma c_j) = l_j \qquad (6)$$

Expressions (4) and (5) are integrated into expression (6), and the histogram statistic f_i are computed. To determine the long-term statistic of the foreground movement effects and create the extracted feature greatly differentiated, the spatial block $\{A_i, A_{i+1}, \dots, A_{i+n-1}\}$with $n$successive frame is taken as a space-time divider$\tilde{A}_i$. In line with expression (6), the character portrayal of every space block in $\tilde{A}_i$ can be computed. The respective weight vector of provided space-time blocks can be articulated as $\{g_i, g_{i+1}, \dots ., g_{i+n-1}\}1 \leq i \leq S \times G$. The feature portrayal of the n-frames space blocks takes average values as running foreground effects features$\tilde{g}_i(1 \leq i \leq S \times G)$ of space-time blocks$\tilde{A}_i$. The computation formulation is

$$\tilde{g}_i = \frac{1}{n}\sum_{j=0}^{n-1} g_{i+1}\ ,\ \tilde{g}_i \in R^q \qquad (7)$$

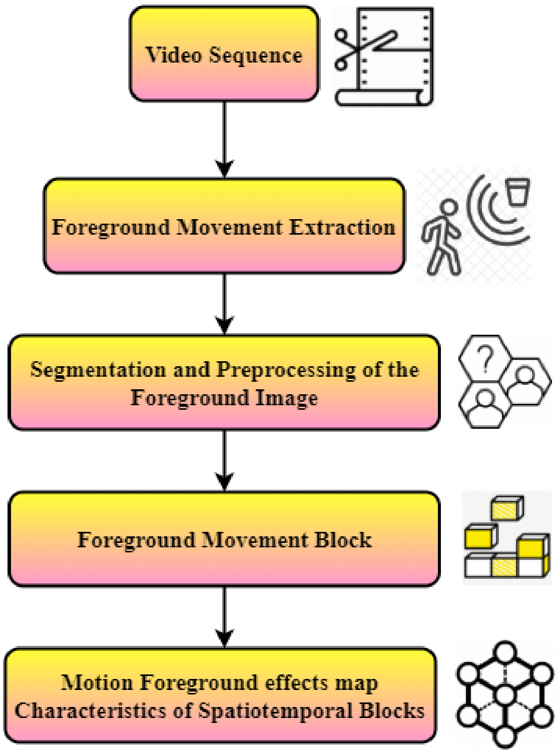
Video Sequence
Foreground Movement Extraction
Segmentation and Preprocessing of the Foreground Image
Foreground Movement Block
Motion Foreground effects map Characteristics of Spatiotemporal Blocks

Figure 3: Flowchart of Feature Extraction

Figure 3 shows the flowchart of feature extraction. Firstly, the adaptive Gaussian mixture model extracts the foreground movement-image from the video frame series and determines spatial blocks. The spatial blocks are joined with the foreground imageries and then pre-processed to determine the foreground movement blocks. The movement illustration of the foreground movement block is determined in line with the dense optical flows of video frames. In conclusion, the effect weight vector for every foreground movement block is computed for every spatial block. The effect weight vector of numerous successive frames of space block means determining a feature depiction of provided space-time blocks.

The weighted sum of squares minimizes the optical flow in a neighborhood:

$$\sum \lambda^2\,(b)(F_b * \beta + F_a * \alpha + F_c) \qquad (8)$$

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

Among them, $\lambda^2(b)$denotes the window weight so that the weight in the neighborhood is higher than the surrounding ones.

The particular stages of the algorithm are:

(i) Firstly, the Gaussian pyramid is implemented for every frame. As the number of pyramid layers rises, the resolutions of the picture progressively reduce.

(ii) Compute optical flow. Analyzeconstantly from the top layer by reducing the errors and computing the optical flows in the top picture:

$$Optical\ flow = \sum\sum\left(F(b,a) - G(b+\omega_b, a+\omega_a)\right)^2 \quad (9)$$

The displacement of every layer

$$\omega_m = \frac{\omega}{2^m}$$

Among them, ωdenotes the displacement sizes of the actual picture and $m$indicates the number of layers.

(iii)Every layer input of the picture is the output of the prior layer; link it to the end, and compute the values of the optical flows of every layer in series.

$$Optical\ flow_{u_{m-1}} = 2 * optical\ flow_{u_m} + \omega_m \quad (10)$$

(iv)It can be seen that the attainment of the optical flow values is a superposition of the optical flow vector of every layer.

(v) Compute from top to bottom along with pyramids, repeat the approximation operations, and acquire the size of the optical flows of the bottom picture.

$$Optical_u = \left(optical\ flow_{u_0} + \omega_0\right) \quad (11)$$

The benefit of the pyramid Lucas-Kanade optical flow technique is that the optical flow value computed every time is comparatively less. The outcome is to intensify every value additively. In this manner, lesser neighborhood windows can manage greater pixel movements.

2D CNN with a network architecture is mainly used in this field. The original 2D architecture was used for temporal information to increase accuracy. Additionally, it has been discovered that 3D spatiotemporal convolutions effectively identify actions in videos.

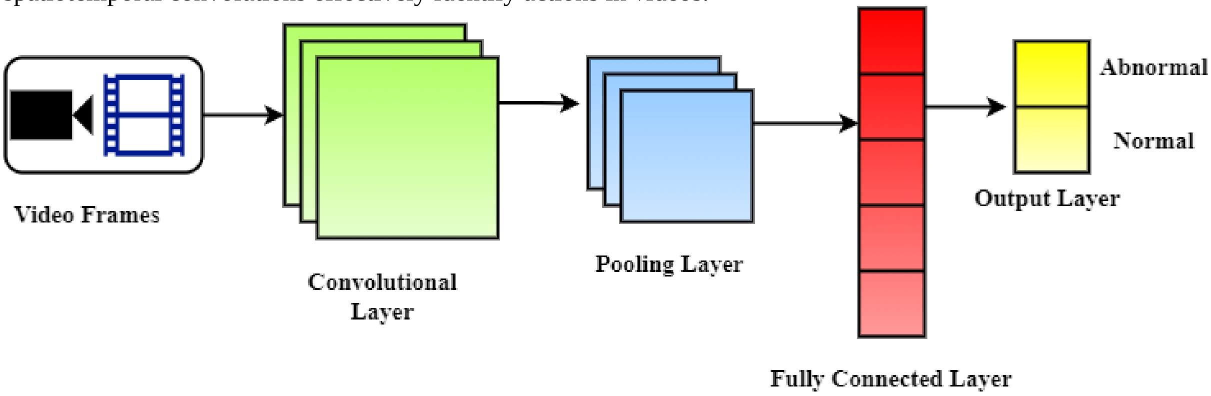

Figure 4: CNN Model

The CNN model is displayed in Figure 4. The use of pre-trained CNNs made for related domains can enhance the performance of aberrant behavior recognition. Thus, it is expected that a method that extends the idea by fine-tuning a video-pre-trained 3D network first trained for general behavior categorization to enable anomalous behavior detection will improve the recognition performance. First, a video dataset is painstakingly put together by selecting movies from publically accessible datasets with consideration for the underlying logic of each reported analyzed behavior. After that, these movies undergo pre-processing to create a format suitable for CNN input. Second, a two-stream CNN architecture improves recognition performance by combining video (RGB) and optical flow streams. Third, the absence of instances in public video sets affects the ability to recognize deviant behavior.

The cross-entropy loss function K as given below

$K(x,t) = -\sum_{D=1}^{D} t_D.\log(x_D)$ (12)

Equation (12) illustrates this, where D-dimensional vector x represents the prediction of the network as a vector of 0 and 1, and D-dimensional vector t represents the ground truth values. One class's

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

significance in this equation can be raised by including weight in the loss function. As a result, the following provides our modified loss function L:

$$K(x,t) = -\sum_{D=1}^{D} \omega_D . t_D . \log(x_D) \qquad (13)$$

As inferred from equation (13), where ω are weights, t is the ground truth, and x is the prediction.

In this method, there is no need to change the weighting of the class as assigned as 1.0. Every error in a class with a larger weight, or the loss function, will be penalized more severely than any other class. The network prioritizes the learning of some classes by raising their weight. The current application assessed the class weights by $compute\_class\_weight$ in $scikit\_learn$ and charity the parameter $class\_weight$ though exercise the classical.

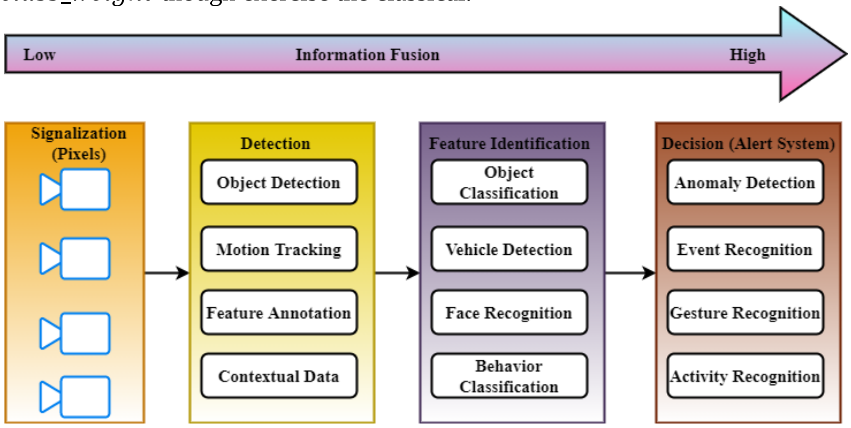

Figure 5: Processing tasks of video analytics

Figure 5 shows the processing tasks of video analytics. Only the image that includes relevant data (such as an object identified) is processed using Measurement Level Fusion since it deals with loading, decoding, and dimensionality reduction of incoming data streams. To determine the route of measure and the properties of an object, like size, form, colors, etc., a CNN model for object detection is required (feature annotation). The image data is combined with contextual data like date stamps, geolocation, and camera detection at this level, which takes place in the cloud. To fully comprehend the scene that has been recorded, it is critical to combine all of the available information. Although this stage of processing aids in further inference and decision-making, it also acts as a filter to lower the volume of data delivered and processed, hence lowering latency, bandwidth, and energy consumption. The feature level fusion identifies the object or event of interest in the imageries captured in the previous phase, corresponding to the next processing stage. It lowers the quantity of data processed by employing a convolutional neural network model for pattern recognition and extracting particular areas or features from the resulting picture. Pattern recognition is often used to recognize vehicles, pedestrians, and even individuals via facial recognition for smart cities and buildings. After all, finally, at the Decision Level Fusion stage (alert system). A global view of actions and events may be inferred from various lower-level abstraction data observations. As an example of a complicated phenomenon, this processing step may evaluate and take appropriate action when someone attempts to enter an area where they are not allowed. Before choosing, this study gathers more details about the scenario and derives useful knowledge from the identified characteristics.

3. Simulation Analysis and Discussion

The recommended ML-IVSS model's experimental findings have been implemented based on performance metrics like recall ratio, abnormal behavior detection ratio, prediction ratio, accuracy ratio, precision ratio, F1-score ratio, and loss value comparison.

(i) Abnormal Behavior Detection Ratio

With the growing number of surveillance cameras, supervising numerous monitors by security agents becomes problematic owing to individual fatigue and inattention. In addition, abnormal events are comparatively rare and do not occur regularly. Thus, increasing demand for a smart video surveillance model automatically identifies abnormal behavior and raises the alarm. The article suggests it augmented spatial-temporal convolutional neural networks to overcome the issues with existing crowd anomalous detection representations, the inability of traditional neural networks to

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

extract time-related features, and the lack of a training sample. Figure 6 demonstrates the abnormal behavior detection ratio. The test stage identifies if the super-pixels are irregular depending on the least rebuilding errors among the super-pixel features. The dictionary blend sets in the test sets are identified.

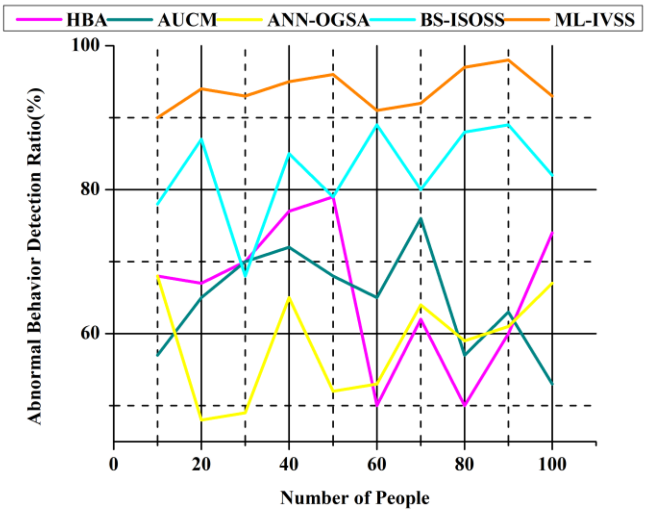

Figure 6: Abnormal Behavior Detection Ratio

(ii) Prediction Ratio

An approach for categorizing every abnormal behavior is assumed. Pedestrian recognition and tracking technique are utilized in a module for determining the pedestrian location, and loitering, or intruding pedestrians in a particular region can be identified utilizing the coordinates of the tracked pedestrian. In this module, pedestrian data is stored to examine the unusual behavior of a similar pedestrian, and the pedestrian image is transmitted to the strange behavior analysis module via IP/TCP communication to predict violence or falls. This study likewise inserts the same object detection and inter-module communication subsystem for precise analysis of abnormal behavior. Figure 7 demonstrates the prediction ratio.

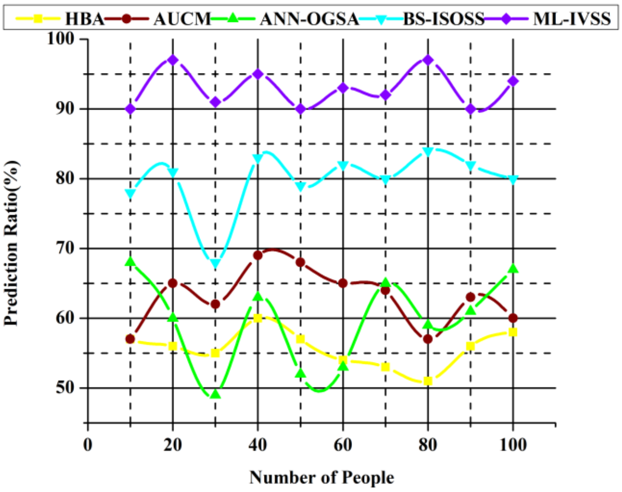

Figure 7: Prediction Ratio

(iii)Accuracy Ratio

Present state-of-the-art computer vision methods still resolve to challenge issues involving rapid motion, light changes, motion blur, occlusions, rotations, and deformation. Besides reliability and accuracy, fast processing is vital to visual surveillance for real-time investigation. Since a single video tracking model cannot resolve the complex irregular behavior analysis issue because of the difficulties mentioned, this study combines object detection and visual tracking approaches

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

recompense for tracking failure and updating re-appearing pedestrians. The accuracy of the model can be computed by

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

As inferred from equation (14), TN, FN, TP, and FP are true negative, false negative, true positive, and false positive, correspondingly. Figure 8 illustrates the accuracy ratio.

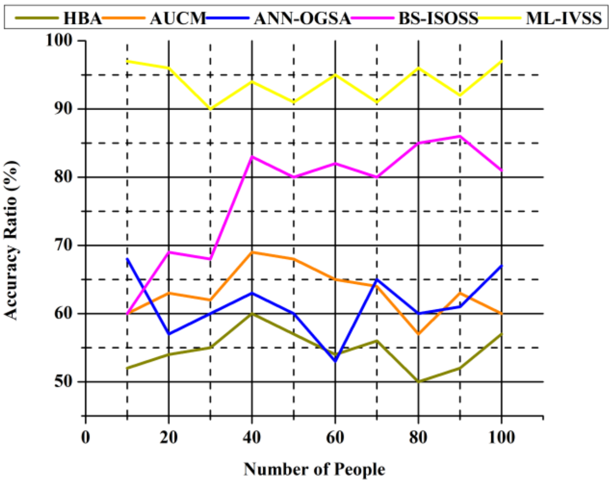

Figure 8: Accuracy Ratio

(iv) F1-Score Ratio

Incorporating all processing stages in single video analytics permits cities to create an autonomous and smart monitoring system. Formerly, individuals performed tasks. They must observe events in every image from cameras distributed around the metropolitan. As well as being inaccurate and tedious for individuals, it was quite costly because it usually convoluted many persons in this monitoring progression. The most popular F-measure is the F1 score, which belongs to a broader parametric class called the F-measures. It can be calculated using the formulas below. It is known as the harmonic mean of recall and precision.

$$F_1 = \frac{2.TP}{2.TP+FP+FN} \quad (15)$$

Figure 9 signifies the F1-score ratio.

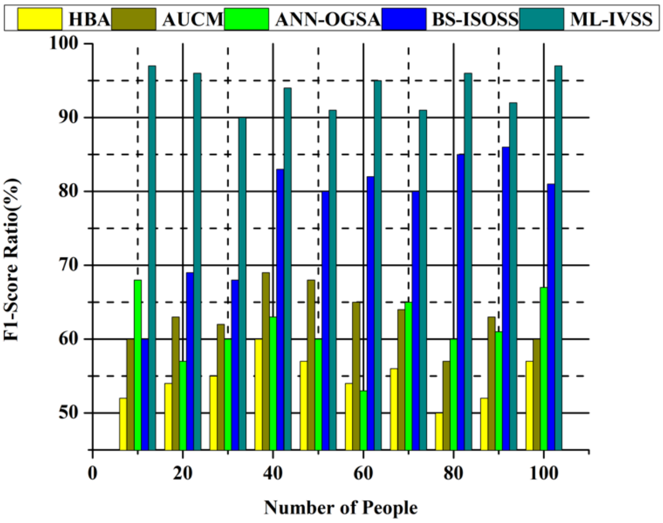

Figure 9: F1-Score Ratio

(v) Recall Ratio

ML approaches have brought important advantages in video/image and sensor data processing, like decreasing data dimensionality and attaining greater recall and precision in the yielded data. Recall, in this instance, gauges the percentage of behaviors accurately anticipated to belong to one of the

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

four classifications. It's referred to as sensitivity. It gauges the percentage of real positives that were accurately identified.

$Recall = \frac{TP}{TP+FN}$ (16)

Figure 10 demonstrates the recall ratio.

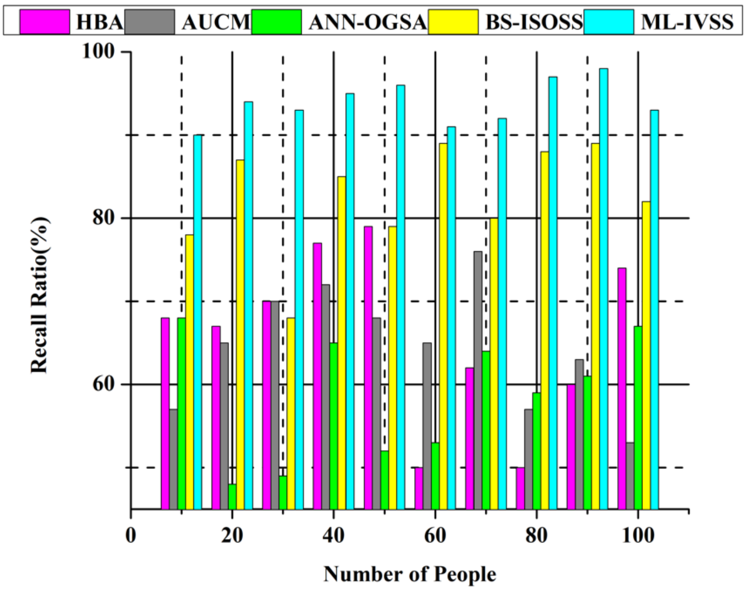

Figure 10: Recall Ratio

(vi) Precision Ratio

The study used two key metrics—precision and recall—since accuracy is not typically regarded as the best indicator of a model's effectiveness because they are unaffected by unequal class distributions. The percentage of valid class identifications is what determines precision.

$Precision = \frac{TP}{TP+FP}$ (17)

As shown in equation (17), TP and FP are true positive and false positive, respectively. Figure 11 shows the precision ratio.

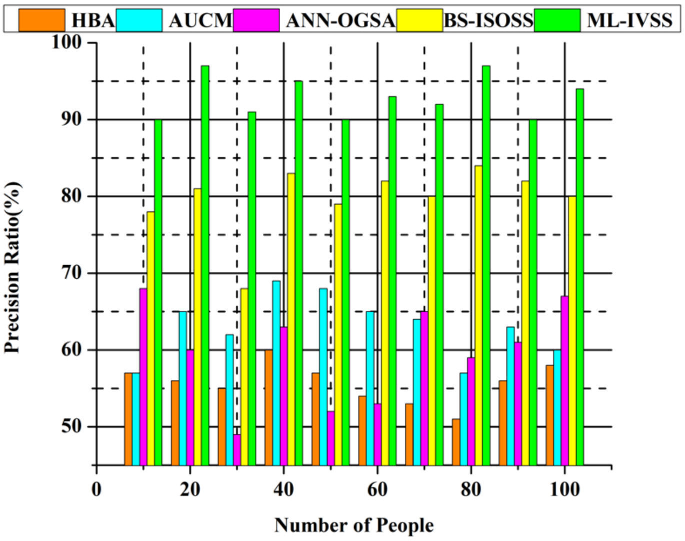

Figure 11: Precision Ratio

Compared to other existing HBA, AUCM, ANN-OGSA, BS, the proposed ML-IVSS model improves the abnormal behavior detection ratio, precision ratio, accuracy ratio, F1-score ratio, prediction ratio, and recall ratio.

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

4. Conclusion

This paper present ML-IVSS model for abnormal behavior detection in smart city surveillance systems, presented in this paper, demonstrates the potential of intelligent fusion techniques for improving detection accuracy while reducing computation costs. By restraining feature extraction to the space-time volume around trajectories, the method utilizes spatiotemporal data and motion information more efficiently, resulting in better performance. Overall, this approach highlights the benefits of intelligent fusion techniques for optimizing the performance of surveillance systems in smart city environments. The essential assumption of this research is that personal conduct is made of a sequence of static postures, and their temporal link facilitates its identification. Comparing the proposed ML-IVSS model to other widely used approaches, the simulation study shows that it improves the abnormal behavior detection ratio of 98.8%, prediction ratio of 97.4%, accuracy ratio of 96.9%, F1-score ratio of 97.1%, the precision ratio of 95.6%, and recall ratio of 96.2%.

Funding: “This research received no external funding”
Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Li, H., Xiezhang, T., Yang, C., Deng, L., & Yi, P. (2021). Secure Video Surveillance Framework in Smart City. Sensors, 21(13), 4419.

[2] Saeed M. Aljaberi , Ahmed N. Al-Masri, Automated Deep Learning based Video Summarization Approach for Forest Fire Detection, Journal of Intelligent Systems and Internet of Things, Vol. 5 , No. 2 , (2021) : 54-61 (Doi : https://doi.org/10.54216/JISIoT.050201)

[3] Shorfuzzaman, M., Hossain, M. S., & Alhamid, M. F. (2021). Towards the sustainable development of smart cities through mass video surveillance: A response to the COVID-19 pandemic. Sustainable cities and society, 64, 102582.

[4] Masud, M., Muhammad, G., Alhumyani, H., Alshamrani, S.S., Cheikhrouhou, O., Ibrahim, S. and Hossain, M.S., 2020. Deep learning-based intelligent face recognition in IoT-cloud environment. *Computer Communications*, *152*, pp.215-222.

[5] Ramprasad, L., & Amudha, G. (2014, February). Spammer detection and tagging based user generated video search system—A survey. In International Conference on Information Communication and Embedded Systems (ICICES2014) (pp. 1-5). IEEE.

[6] Janakiramaiah, B., Kalyani, G., & Jayalakshmi, A. (2021). Automatic alert generation in a surveillance systems for smart city environment using deep learning algorithm. Evolutionary Intelligence, 14(2), 635-642.

[7] Amudha, G. (2021). Dilated Transaction Access and Retrieval: Improving the Information Retrieval of Blockchain-Assimilated Internet of Things Transactions. Wireless Personal Communications, 1-21.

[8] Gheisari, M., Najafabadi, H. E., Alzubi, J. A., Gao, J., Wang, G., Abbasi, A. A., & Castiglione, A. (2021). OBPP: An ontology-based framework for privacy-preserving in IoT-based smart city. Future Generation Computer Systems, 123, 1-13.

[9] Jin, Y., Qian, Z., & Yang, W. (2020). UAV cluster-based video surveillance system optimization in heterogeneous communication of smart cities. IEEE Access, 8, 55654-55664.

[10] Gao, J., Wang, H., & Shen, H. (2020). Task failure prediction in cloud data centers using deep learning. IEEE Transactions on Services Computing.

[11] Do, D. T., Van Nguyen, M. S., Nguyen, T. N., Li, X., & Choi, K. (2020). Enabling multiple power beacons for uplink of noma-enabled mobile edge computing in wirelessly powered IOT. IEEE Access, 8, 148892-148905.

[12] Yoon, C. S., Jung, H. S., Park, J. W., Lee, H. G., Yun, C. H., & Lee, Y. W. (2020). A cloud-based UTOPIA smart video surveillance system for smart cities. Applied Sciences, 10(18), 6572.

[13] Do, D. T., Nguyen, T. T. T., Nguyen, T. N., Li, X., & Voznak, M. (2020). Uplink and downlink NOMA transmission using full-duplex UAV. IEEE Access, 8, 164347-164364.

[14] Zahraa Faiz Hussain, & Hind Raad Ibraheem. (2023). Novel Convolutional Neural Networks based Jaya algorithm Approach for Accurate Deepfake Video Detection. Mesopotamian Journal of CyberSecurity, 2023, 35–39. https://doi.org/10.58496/MJCS/2023/007

[15] Nagrath, P., Thakur, N., Jain, R., Saini, D., Sharma, N. and Hemanth, J., 2022. Understanding new age of intelligent video surveillance and deeper analysis on deep learning techniques for object tracking. In *IoT for Sustainable Smart Cities and Society* (pp. 31-63). Cham: Springer International Publishing.

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023

[16] Elhoseny, M. (2020). Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems. Circuits, Systems, and Signal Processing, 39(2), 611-630.

[17] Song, X. H., Wang, H. Q., Venegas-Andraca, S. E., & Abd El-Latif, A. A. (2020). Quantum video encryption based on qubit-planes controlled-XOR operations and improved logistic map. Physica A: Statistical Mechanics and its Applications, 537, 122660.

[18] Yassine, S., Kadry, S., & Sicilia, M. A. (2020). Statistical Profiles of Users' Interactions with Videos in Large Repositories: Mining of Khan Academy Repository. KSII Transactions on Internet and Information Systems (TIIS), 14(5), 2101-2121.

[19] Kumar, S., & Soundrapandiyan, R. (2021). A multi-image hiding technique in dilated video regions based on cooperative game-theoretic approach. Journal of King Saud University-Computer and Information Sciences.

[20] Kumar, P. M., & Seon, H. C. (2021). Internet of things-based digital video intrusion for intelligent monitoring approach. Arabian Journal for Science and Engineering, 1-11.

[21] Angadi, S., & Nandyal, S. (2020). Human identification system based on spatial and temporal features in the video surveillance system. International Journal of Ambient Computing and Intelligence (IJACI), 11(3), 1-21.

[22] Feizi, A. (2020). Hierarchical detection of abnormal behaviors in video surveillance through modeling normal behaviors based on AUC maximization. Soft Computing, 24(14), 10401-10413.

[23] Appathurai, A., Sundarasekar, R., Raja, C., Alex, E. J., Palagan, C. A., & Nithya, A. (2020). An efficient optimal neural network-based moving vehicle detection in traffic video surveillance system. Circuits, Systems, and Signal Processing, 39(2), 734-756.

[24] Thenmozhi, T., & Kalpana, A. M. (2020). Adaptive motion estimation and sequential outline separation based moving object detection in video surveillance system. Microprocessors and Microsystems, 76, 103084.

[25] Moorthy, K., Ali, M.H., Ismail, M.A., Chan, W.H., Mohamad, M.S. and Deris, S., 2019. An Evaluation of Machine Learning Algorithms for Missing Values Imputation. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(12S2).

[26] Jaber, M.M., Ali, M.H., Abd, S.K., Jassim, M.M., Alkhayyat, A., Alreda, B.A., Alkhuwaylidee, A.R. and Alyousif, S., 2022. A Machine Learning-Based Semantic Pattern Matching Model for Remote Sensing Data Registration. Journal of the Indian Society of Remote Sensing, pp.1-14.

[27] Ali, M.H., Al-Jawaheri, K., Adnan, M.M., Waheed, S.R., Kadhim, K.A. and Rahim, M.S.M., 2021, September. Review of Intrusion Detection Systems Based on Machine Learning. In 2021 4th International Iraqi Conference on Engineering Technology and Their Applications (IICETA) (pp. 195-200). IEEE.

Doi: https://doi.org/10.54216/FPA.110203
Received: November 22, 2022 Accepted: April 09, 2023