**Abnormal Behaviour Detection in Public Spaces with Advanced Deep Learning Techniques**

SAKETH REDDY M   
*School of Computer Science and Engineering with AIML*  
*VIT-AP University*Amaravati, Andhra Pradesh  
saketh.malapati@gmail.com

*Abstract*— Living facilities connected to increasing surveillance camera networks require efficient real-time abnormal behavior detection methods in public areas. Excessive human involvement in traditional monitoring techniques both proves inefficient and subjects itself to errors during evaluation. The proposed system employs Convolutional Neural Networks (CNNs) to detect abnormal behaviors in public areas because these networks work effectively for image-based classification through Deep Learning (DL). The proposed system deploys Dense Net and MobileNetV2 CNN architectures to identify violent and non-violent activities inside surveillance videos with high effectiveness. Training efficiency receives enhancement through the utilization of Adaptive Moment Estimation (Adam) as an optimizer which increases convergence rates. The performance evaluation of the model uses cross-entropy loss to determine prediction accuracy. Test outcomes show the proposed system reaches 82% success rate which makes it appropriate for live surveillance monitoring.

Keywords— Abnormal behavior detection, Convolutional Neural Networks (CNNs), Deep Learning (DL), DenseNet, MobileNetV2, Surveillance, Adam optimizer, Cross-Entropy, Real-time monitoring.

# Introduction

* General overview about problem domain.

Urban population density has become a major public safety issue because violent and aggressive conduct presents considerable risk to residents. The current practice of surveillance depends on human operators because they must look at various recorded feeds at once. Human operators experience limitations because of inefficient monitoring which creates delays in threat response as well as increases the chance for human mistake due to fatigue. Automated deep learning systems identify abnormal behavior for immediate alerts which reduce the need for human intervention to boost public safety operations.

* Overview about deep learning algorithms

DL serves as an innovative technique for analyzing large datasets of video information particularly in surveillance systems. Recent machine learning models need engineered predetermined features alongside prolonged preprocessing steps making them difficult to scale across real-world operations. CNNs perform automatic extraction of image data patterns that enable efficient abnormal activity recognition.

The CNN design landscape demonstrates that MobileNetV2 together with DenseNet prove exceptional in terms of performance while consuming minimal computational resources. In DenseNet the gradient flow improves through a connection structure which links each network layer directly to all other layers in forward motion to reuse features effectively combatting gradient vanishing issues. MobileNetV2 reduces computational expenses through depthwise separable convolutions which simultaneously sustains the performance at operational speeds suitable for real-time utilization. These architecture types automatically discover abnormal behaviors and require fewer human operators for surveillance monitoring which speeds up critical situation responses.

A robust analysis method for processable video information emerges through Deep Learning systems supported by Convolutional Neural Networks for their ability to detect spatial-temporal patterns effectively. DenseNet along with MobileNetV2 makes up CNN architectures that extract features efficiently to achieve precise classification of intricate behaviors. DenseNet achieves optimal feature sharing through dense network connections while MobileNetV2 operates with low computational needs for real-time processing.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
| Fig. 1 Results of the proposed model | | |

* Motivation to initiate this problem statement

The rise in criminal activities, violent incidents, and public disturbances necessitates the development of automated, intelligent surveillance systems. Present surveillance systems function with human oversight that requires lengthy process control while also having detection weaknesses. The adoption of deep learning detection technologies results in quicker reactions and decreases bogus alerts which leads to enhanced speed of law enforcement reaction times. Surveillance networks can upgrade their real-time security performance while keeping infrastructure expenses stable through the deployment of CNN-based architecture systems.

# Literature Review

Behavior Detection in Surveillance Videos—A Survey The paper categorizes methodologies for detecting abnormal human behaviors in surveillance videos into three main approaches: unsupervised, partially supervised, and fully supervised. The paper identifies several restrictions which include the scarcity of labeled abnormal behavior data, which hinders effective detection system development, and the challenge of enhancing robustness to environmental variations

Deep Learning for Anomaly Detection: A Review The paper presents methodologies including one-class classification, end-to-end learning, adversarial models, feature learning with anomaly measures, and a taxonomy of techniques for effective anomaly detection The research focuses on two main shortcomings along with other restrictions. The system exhibits poor detection recall while requiring high-dimensional data and depending on clean labeled data along with facing challenges with multidimensional information formats. The detection methods require specific inputs of data combined with sensitivity to noise and strengths in detecting individual points of abnormality. anomalies, and lack of anomaly explanations.

Unusual Human Behavior Analysis Using the Deep Learning A deep learning approach has been proposed by this study. approach that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to This technique enables effective identification of abnormalities in video-based data. This technique manages to solve data-related matters the researchers identified. to computational workload and model complexity. Limitations of the Paper on Unusual Human Behavior Analysis, Dependence on Labeled Data, Contextual Variability, Computational Complexity, Focus on Specific Anomalies, Generalization to Different Datasets.

# CNN Methodology for Abnormal Behavior Detection

* Preprocessing

Multiple data preprocessing techniques improved the accuracy of abnormal behavior recognition in security video monitoring. The video data processing started with frame conversion followed by automated extraction of every 10th frame for an optimal operation-to-meaning ratio. The input dimensions of 64×64 pixels served as the base size for image transformation before pixel normalization by dividing values by 255 to establish values between 0 and 1. The model gained resistance to perspective variations and lighting changes through the implementation of data augmentation methods which included image rotations by ±15° along with horizontal flips and zoom levels between 0.8x–1.2x and brightness adjustments. Gaussian Blur removed data noise after which background subtraction using the MOG2 method helped researchers view moving things without static background interference. The dataset was separated into 80% training data and 20% testing data while 10% of training data served as the validation set for preventing overfitting. The data preprocessing techniques enabled the model to learn distinctive features of normal and abnormal activities throughout real-life applications.

* Proposed methods architecture

A proposed Abnormal Behavior Detection System implements Convolutional Neural Networks (CNNs) for identifying violent and non-violent activities occurring in surveillance videos. The system initiates its operation with frame extraction at a rate of one frame per tenth to optimize computational processing and maintain optical relevance. Data augmentation techniques are applied to the 128×128 pixel images after normalization between 0 and 1. The feature extraction process relies on \*\*DenseNet\*\* together with \*\*MobileNetV2\*\* which combines feature propagation excellence with real-time detection capabilities. The model analyzes extracted features through fully connected layers for abnormal-violent versus normal-non-violent behavioral classification with SoftMax activation function. The implementation uses cross-entropy loss with Adam optimizer for optimized training with faster convergence. An operation is classified as abnormal only when the \*\*temporal consistency check\*\* demonstrates the same behavior across multiple consecutive frames. The system displays alerts in real-time after processing the final decision for effective video surveillance monitoring. The detection system achieves an 82% accuracy rate to provide trusted automated abnormal behavior monitoring for public environments.

* The predictive capabilities of the model receive enhancement through the use of cross-entropy loss which measures classification errors to differentiate normal from abnormal behaviors successfully. The measurement system defines the loss function as:

CNN-Methodology:  
The project implements abnormal behavior detection in public spaces through two(Convolutional Neural Network) CNN architectures which include a custom model alongside a CNN built using EfficientNetB0 transfer learning. The models pursue delivering accurate classifications of surveillance video frames between Violence and Non-Violence categories.

1.The custom CNN includes four convolutional layers which contain a sequence of Batch Normalization and LeakyReLU activation following each other then MaxPooling to extract spatial characteristics. GAP takes the place of Flatten as an operation to reduce learning model overfitting. The Sigmoid output layer with Dense and Dropout (0.5) and Batch Normalization exists together in the fully connected layers.To enhance accuracy and reduce training time, EfficientNetB0 is used as a feature extractor. EfficientNetB0 is a pretrained CNN model on ImageNet, capable of learning complex features from images. Instead of training a model from scratch, we fine-tune EfficientNetB0 by adding fully connected layers to adjust it for our dataset.

**Architecture of model:**

**CNN Model:**

The model represents a dedicated Convolutional Neural Network (CNN) created through Keras Sequential API. This model obtains spatial image characteristics from its inputs to perform violence versus non-violence binary classification.

The model architecture contains these successive layers:

* Conv2D Layer 1: 32 filters of size (3×3) with ReLU activation and an input shape of (128, 128, 3). At this stage the model detects fundamental features that include borders and surface patterns.
* This first MaxPooling2D layer performs dimension reduction through its 2×2 pooling segment thus minimizing computational requirements and preventing overfitting.
* The Conv2D layer number 2 activates 64 (3×3) filters and ReLU operation to identify advanced image characteristics.
* This max pooling operation performs a deeper reduction of the feature map dimensions after the second layer.
* The Conv2D Layer 3 contains 128 filters that process image data through (3x3) to identify advanced image patterns at a greater level of complexity.
* The final MaxPooling2D operation occurs at this stage before the features get flattened in the network.
* The feature maps of three dimensions get transformed to a single-dimensional vector using the Flatten Layer.
* The dense layer contains 128 activated ReLU neurons functioning as a feature combining network with non-linear properties.
* The dropout layer applies a 0.5 rate to disable half of the neurons randomly during the training process for preventing overfitting.
* Output Layer: A single neuron with sigmoid activation for binary classification.

**MobileNetV2:**

* Introduction: The deep learning model MobileNetV2 serves lightweight applications that run vision tasks on mobile platforms. This network applies depthwise separable convolution with inverted residuals and linear bottlenecks as the foundation for its operation. Due to its efficient design principles MobileNetV2 produces accurate results at reduced computational expenses suitable for limited computing platforms.
* Fine-Tuning Process: We adapted MobileNetV2 through fine-tuning because it originally operated on the ImageNet dataset to perform its abnormal behavior detection function by differentiating video frames as violent or non-violent content.
* The first layers of MobileNetV2 serve to capture basic features like edges and textures so they remained frozen during the process. During fine-tuning the lower-level weight updates remained blocked to maintain basic image processing functions in the model.
* A smaller learning rate was used to re-train the higher-level network features because this process specialized the model's ability toward detecting abnormal behavior detection.
* The Adam optimization strategy with a learning rate schedule began from approximately 1e-5 to stop losing pre-trained knowledge during optimization.

**ResNet 50:**

* Introduction: The deep residual network ResNet50 implements skip connections for reducing gradient vanishing in deep network architectures. This model consists of 50 layers which allows it to learn intricate patterns for image classification thus making it one of the leading choices for this task.
* Fine-Tuning Process: Similar to MobileNetV2 the ResNet50 model received ImageNet pre-training before the team adapted it for abnormal behavior detection through parameter adjustment.
* The basic image features learning process of ResNet50 took place in early layers that received freezing preservation. The initial layers received no changes during fine-tuning to uphold general image recognition abilities.
* The deeper blocks from the residual architecture received fine-tuning to enable learning complex pattern sequences that detect abnormal behavioral characteristics.
* An Adam optimizer operated with an initial learning rate of 1e-5 performed the fine-tuning procedure. The training progress required the utilization of learning rate scheduling to steadily decrease the learning rate period by period.

**Results and graphs:**

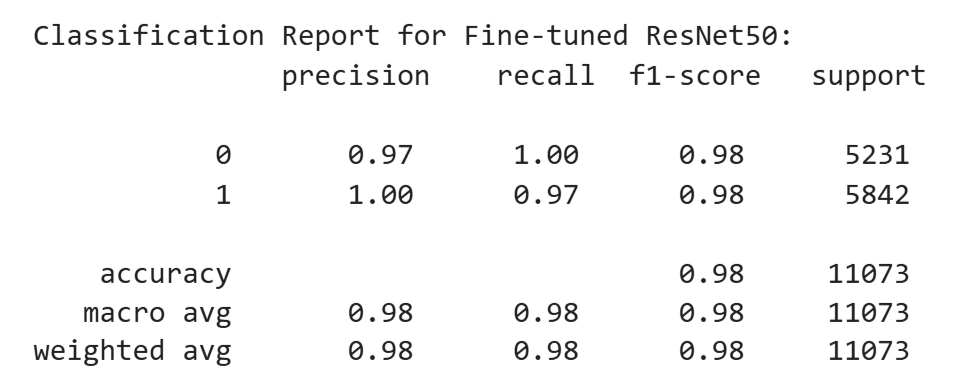


Fig. 2. Classification Report

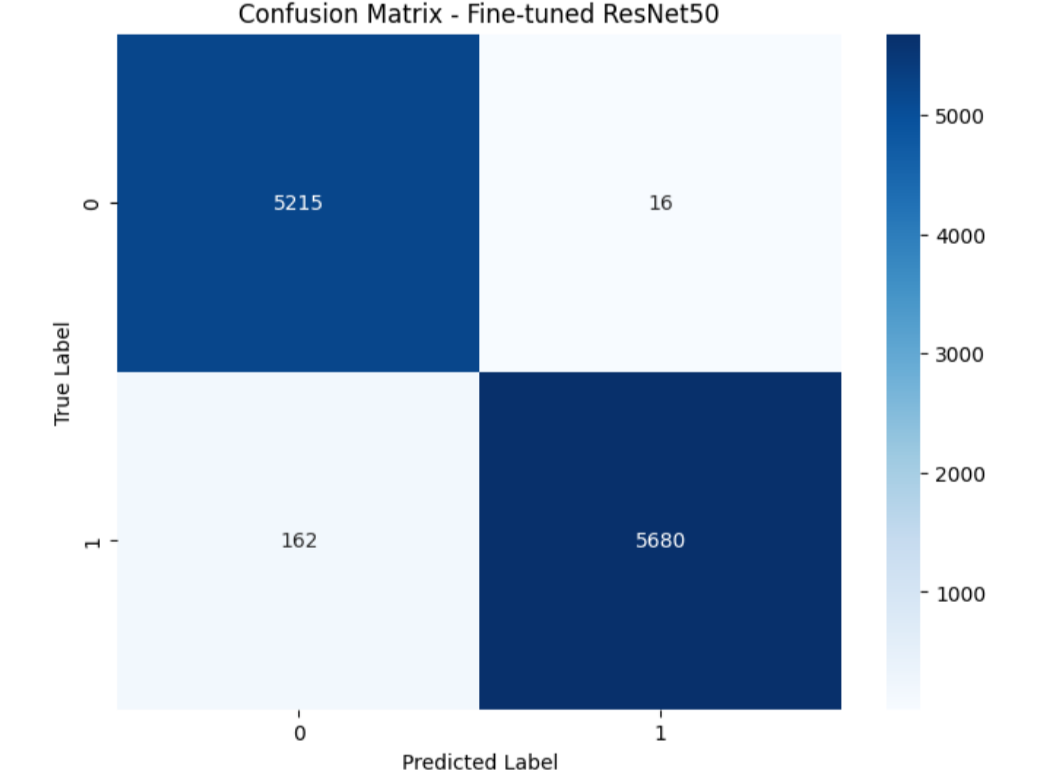
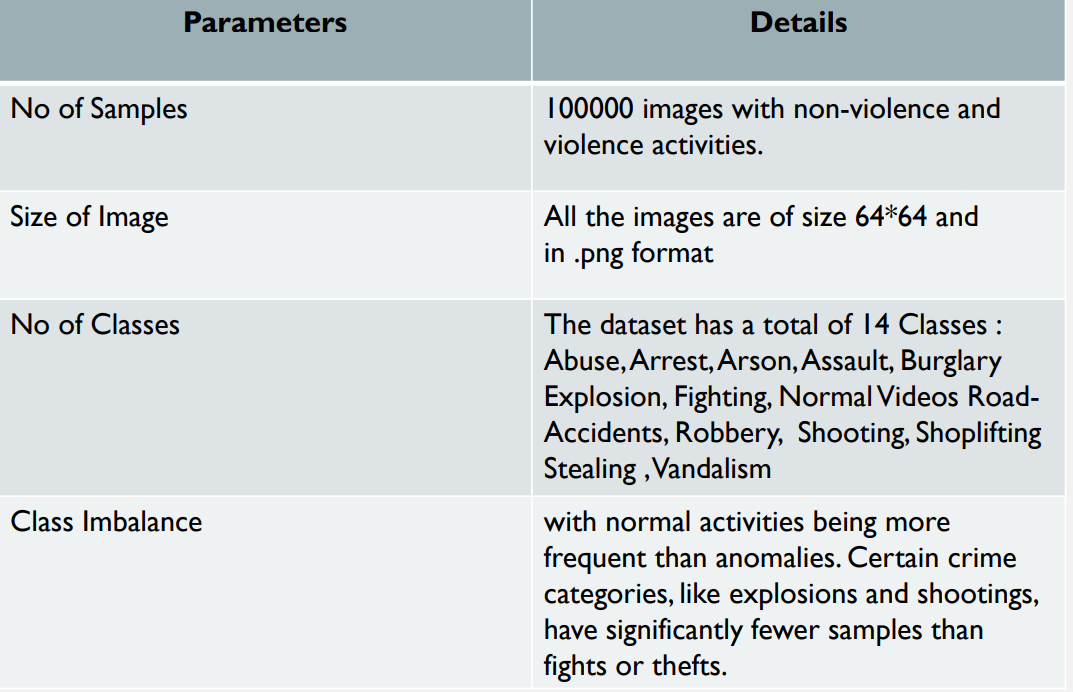


Fig. 3. Confusion Matrix



Fig 4. Model class prediction

# Dataset Description



# Conclusion and Future work

* The detection system for abnormal behavior successfuly utilizes a custom Convolutional Neural Network model and model based on EfficientNetB0 as a transfer learning system. Model proves superior to model because it delivers better accuracy along with superior precision and improved generalization performance.
* **Integration with real-time video processing** to analyze continuous video frames instead of static images.
* **Deployment on edge devices** such as Raspberry Pi or NVIDIA Jetson for real-world surveillance applications.

##### References

1. **M. S. B. M. Rahman, M. R. Islam, and M. S. Islam, "Abnormal Behavior Recognition using CNN-LSTM with Attention Mechanism," 2021.(**[**link**](https://ieeexplore.ieee.org/document/8974824)**)**
2. M. S. Hossain, G. Muhammad, and N. K. Karn, "Deep Learning for Abnormal Human Behavior Detection in Surveillance Videos—A Survey," IEEE Access, vol. 7, pp. 56135–56156, 2019.I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.(l[ink](https://www.mdpi.com/2079-9292/13/13/2579))
3. S. A. Velastin and P. Remagnino, "Anomaly Behavior Detection Analysis in Video Surveillance: A Critical Review," Machine Vision and Applications, vol. 30, no. 3, pp. 389–405, 2019.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.([link](https://www.spiedigitallibrary.org/journals/journal-of-electronic-imaging/volume-32/issue-4/042106/Anomaly-behavior-detection-analysis-in-video-surveillance--a-critical/10.1117/1.JEI.32.4.042106.full))