Turkish Journal of Security Studies

ISSN: 2148-6166 e-ISSN: 2757-7716
Volume: 28 Article Type: Research
Received: 01.01.2026 Accepted: 14.04.2026

Comparative Evaluation of CNN Architectures for Frame-Level Behavior Recognition in Private Security Surveillance

Ümit YILMAZ[*]

Abstract: The integration of intelligent video surveillance systems has significantly strengthened the role of deep learning in improving private security. In this study, three convolutional neural network (CNN) architectures, EfficientNet-B0, ResNet50, and MobileNetV2, were evaluated for a multi-class behavior recognition task in the context of private security. Behaviors such as abuse, burglary, and shoplifting were detected from surveillance video data by utilizing a dataset comprising ten distinct classes, nine illicit behaviors, and one representing normal activity. Transfer learning techniques were utilized to adapt the CNN models for this classification task. Model performances were assessed based on precision, recall, F1-score, confusion matrices, and multi-class ROC curves. ResNet50 was determined to achieve the highest performance with 99.08% accuracy on the test set, outperforming MobileNetV2 (98.78%) and EfficientNet-B0 (98.72%). It was determined that all three models achieved AUC scores of 1.00 across all behavior classes, indicating high discriminative capability. It was concluded that modern CNN architectures provide effective tools for enhancing surveillance analytics in private security and have significant potential for early threat detection and crime prevention.

Keywords: Private Security, Video Surveillance, Behavior Recognition, EfficientNet, ResNet, MobileNet

[*] Instructor Dr., Bursa Technical University, Quality Coordination Office, umit.yilmaz@btu.edu.tr, ORCID ID: 0000-0003-4268-8598

Güvenlik Çalışmaları Dergisi

ISSN: 2148-6166 e-ISSN: 2757-7716
Cilt: 28 Makale Türü: Araştırma
Makalenin Geliş Tarihi: 01.01.2026 Makalenin Kabul Tarihi: 14.04.2026

Özel Güvenlik Gözetiminde Kare Düzeyinde Davranış Tanıma için CNN Mimarilerinin Karşılaştırmalı Değerlendirmesi

Ümit YILMAZ[*]

Öz: Akıllı video gözetim sistemlerinin entegrasyonu, derin öğrenmenin özel güvenliğin geliştirilmesindeki rolünü önemli ölçüde güçlendirmiştir. Bu çalışmada, EfficientNet-B0, ResNet50 ve MobileNetV2 olmak üzere üç evrişimsel sinir ağı (CNN) mimarisi, özel güvenlikle ilgili çok sınıflı bir davranış tanıma görevi bağlamında değerlendirilmiştir. Dokuz yasadışı davranış ve bir normal aktiviteyi temsil eden on farklı sınıftan oluşan bir veri kümesi kullanılarak gözetim videosu verilerinden kötüye kullanım, hırsızlık ve mağaza hırsızlığı gibi davranışlar tespit edilmiştir. CNN modellerinin bu sınıflandırma görevine uyarlanması için transfer öğrenme tekniklerinden yararlanılmıştır. Model performansları; kesinlik, duyarlılık, F1-skoru, karışıklık matrisleri ve çok sınıflı ROC eğrilerine dayalı olarak değerlendirilmiştir. Test kümesinde ResNet50'nin %99,08 doğruluk ile en yüksek performansı gösterdiği belirlenmiş; MobileNetV2 (%98,78) ve EfficientNet-B0 (%98,72) modellerine kıyasla üstün performans sergilediği tespit edilmiştir. Tüm davranış sınıflarında üç modelin de 1,00 düzeyinde mükemmel AUC skorlarına ulaştığı belirlenmiş; bu bulgu modellerinin yüksek ayırt edicilik kapasitesine sahip olduğu görülmüştür. Modern CNN mimarilerinin, özel güvenlikte gözetim analizlerini geliştirmek için etkili araçlar sunduğu ve erken tehdit tespiti ile suçun önlenmesi açısından önemli bir potansiyel taşıdığı sonucuna varılmıştır.

Anahtar Kelimeler: Özel Güvenlik, Video Gözetimi, Davranış Tanıma, EfficientNet, ResNet, MobileNet

[*] Öğr. Gör. Dr., Bursa Teknik Üniversitesi, Kalite Koordinatörlüğü, umit.yilmaz@btu.edu.tr, ORCID ID: 0000-0003-4268-8598

Introduction

In private security operations, there has been increasing reliance on advanced technology, particularly artificial intelligence (AI), to enhance monitoring and threat detection. Intelligent video surveillance systems, capable of automatically recognizing abnormal or illicit human behaviors, have been regarded as essential for ensuring safety in both public and private environments (Saket et al., 2025). Actions such as fighting, theft, or vandalism can now be detected in real time, enabling security personnel to respond swiftly and proactively prevent incidents. In this context, deep learning has been recognized as a transformative approach that supports real-time analysis of visual data, allowing modern AI-powered surveillance systems to surpass the limitations of traditional human-monitored CCTV systems. These developments have emphasized the strategic role of deep neural networks in proactive crime prevention within private security surveillance.

A growing body of research has focused on automatic behavior recognition and anomaly detection in surveillance videos. Early approaches, based on handcrafted features and simple motion heuristics, were found to lack robustness in complex environments (Khanam et al., 2024). With the emergence of deep learning, especially CNNs, significant gains have been reported by learning hierarchical visual features directly from raw video data (Duong et al., 2023). CNN-based architectures have been successfully deployed in detecting various threat behaviors such as violence, shoplifting, and burglary (Ansari & Singh, 2022; Saket et al., 2025).

A wide range of AI models and hybrid frameworks has been introduced in recent studies for security-focused behavior recognition. For instance, Amin et al. (2024) enhanced the EfficientNet-B0 architecture with a Convolutional Block Attention Module (CBAM) to detect public health violations, achieving a notable increase in accuracy from 87% to 96%. Mohanaprakash et al. (2024) proposed a YOLOv5–Conv2DNet framework for identifying abnormal activities such as loitering and fighting, while Ahmed and Naib (2023) reviewed traditional and modern approaches to object detection, classification, and tracking in surveillance settings.

Hybrid techniques, in which handcrafted features are combined with CNNs, have also been explored. Bhimavarapu et al. (2023) improved classification performance by incorporating SLUP-based CNN weight optimization. Jaggi et al. (2023) integrated YOLOv7 with deep CNNs for real-time detection in crowded environments, addressing limitations of conventional CCTV monitoring. Kerachi et al. (2023) further optimized inference time for embedded surveillance by evaluating combinations of MobileNetV2, EfficientNetB0, and RNNs.

The push for lightweight yet accurate models has been evident in recent literature. Ngoc et al. (2023) developed a 3DCNN model tailored for edge

devices, while Akshith et al. (2022) built a real-time activity recognition platform using LRCN and ConvLSTM. Transfer learning has been leveraged effectively, as shown by Alamuru and Jain (2022), where pretrained models outperformed custom CNNs. Additionally, Lu et al. (2022) achieved high-speed performance through a multi-branch CNN-GRU structure. It has been confirmed by studies such as those of Razak et al. (2022) and Silva et al. (2022) that even compact CNN architectures can achieve high recognition accuracy when appropriately optimized and trained.

A notable gap in the existing literature concerns the limited use of spatio-temporal modeling for video-based anomaly detection. Frame-level CNN analysis, while computationally efficient, does not explicitly capture motion dynamics or temporal dependencies between successive frames. Recent studies have demonstrated the advantages of architectures that incorporate temporal context, such as 3D CNNs, CNN+LSTM hybrids, ConvLSTM networks, and Transformer-based models. For instance, Elmetwally et al. (2025) proposed a deep learning framework for real-time video anomaly detection that leverages temporal feature learning, while Pathirannahalage et al. (2025) conducted a comprehensive analysis of real-time video anomaly detection methods for both human and vehicular movement. Furthermore, Natha et al. (2025) introduced a deep BiLSTM attention model that jointly captures spatial and temporal anomalies in surveillance video. These works collectively highlight the importance of temporal context in building robust surveillance systems and serve as key references for situating the present study's contributions and limitations.

Despite these advancements, most existing studies have focused on a single abnormal behavior or relied on a single CNN model. Comparative evaluations across multiple CNN architectures on a broader set of private security-relevant behaviors remain scarce. To address this gap, the present study offers a comprehensive assessment of three prominent CNN models, EfficientNet-B0, ResNet50, and MobileNetV2, on a multi-class classification task involving ten categories of surveillance video data, encompassing both illicit and normal human behaviors.

The remainder of this paper has been organized as follows. In the Methodology section, the dataset, preprocessing steps, CNN models, training strategy, and evaluation metrics are described. The Results section presents the performance outcomes for EfficientNet-B0, ResNet50, and MobileNetV2, including per-class metrics, confusion matrices, and ROC curves. These results are interpreted in the Discussion section, with an emphasis on performance differences and implications for AI deployment in private security. Finally, the Conclusion section summarizes the key contributions and suggests directions for future research. Through this comparative analysis, practical guidance is aimed at providing for selecting AI models to strengthen private security via automated surveillance.

Methodology

Dataset and Preprocessing

In this study, a multi-class surveillance dataset was constructed from publicly available video resources to recognize behaviors relevant to private security operations. The dataset was derived from the Kaggle platform under the title "Real-Time Anomaly Detection in CCTV Surveillance", which provides a structured version of the UCF-Crime dataset (Kaggle, 2025). The original dataset contains approximately 1,900 real-world CCTV videos across 13 behavior classes and was compiled from YouTube and LiveLeak video searches by 10 annotators with varying levels of expertise in computer vision. The entire dataset comprises 103.8 GB of surveillance footage.

For the purposes of this study, a subset of 10 behavior classes was selected, focusing on scenarios most relevant to private security applications. These classes include: Abuse, Assault, Burglary, Fighting, Robbery, Shoplifting, Stealing, Vandalism, Shooting, and Normal. From each of the selected classes, the first 50 video files were chosen. To generate input data suitable for image-based deep learning models, each video was sampled at regular intervals, and one frame was extracted every 50 frames using OpenCV. As a result, a frame-based dataset was constructed where each extracted image represented a labeled behavior class. All frames were saved in a hierarchical folder structure and labeled automatically according to their parent class.

After frame extraction, a total of 44,740 image samples were obtained, representing a balanced and diverse distribution across the 10 behavior categories. These samples were randomly partitioned into training, validation, and testing sets using a 70–15–15 split, resulting in 31,318 training, 6,711 validation, and 6,711 testing instances. The random split was controlled with a fixed seed to ensure reproducibility.

Before training, a uniform preprocessing pipeline was applied to all images using the torchvision. transforms module. Each image was resized to 160×160 pixels to meet the input requirements of all CNN models employed in this study (EfficientNet-B0, ResNet50, and MobileNetV2). The images were then converted into tensors using standard tensor conversion.

All training, validation, and test images retained their original visual characteristics aside from resizing. Class distributions were observed to be reasonably balanced following the 70–15–15 split. The validation set was used exclusively to monitor early stopping during training; the model weights corresponding to the lowest validation loss were saved and restored upon training termination.

This preprocessing protocol and dataset construction methodology were consistently applied across all CNN architectures evaluated in the study.

CNN Architectures

Three pretrained CNN models, EfficientNet-B0, ResNet50, and MobileNetV2, were fine-tuned in this study, each selected to represent a trade-off between classification accuracy and computational efficiency. All models were initialized with weights pretrained on the ImageNet dataset, a widely adopted practice intended to enhance convergence and generalization, particularly when working with moderately sized datasets. The final classification layers of the networks were modified to output predictions for 10 behavior classes, and a softmax activation function was applied to enable multi-class classification.

ResNet50

ResNet50 has been introduced as a deep convolutional neural network architecture designed to address the vanishing gradient problem, which is commonly encountered in deep neural networks. This issue has been effectively mitigated through the incorporation of residual learning blocks and skip connections, enabling gradient signals to propagate directly across layers and thereby preserving learning performance in deep architectures (Bohlol et al., 2025; Mahjoubi et al., 2025).

The model comprises 50 layers and has been widely employed in image classification, object detection, and segmentation tasks. Despite its considerable depth, ResNet50 has been shown to remain trainable due to its residual structure, which facilitates the learning of complex patterns without suffering from degradation problems often seen in very deep networks (Liao et al., 2025; Rath et al., 2025). Furthermore, when compared with architectures such as VGG, ResNet50 has demonstrated a more efficient use of computational resources, as its design requires fewer floating point operations while maintaining high classification accuracy (Liao et al., 2025).

The architecture was initialized with weights pretrained on the ImageNet dataset, which allowed the model to transfer generic visual features effectively to new domains through fine-tuning, a strategy especially beneficial in tasks involving limited data, such as medical diagnosis and surveillance (Rath et al., 2025). Moreover, by leveraging residual modules, ResNet50 has been shown to improve both inference speed and classification accuracy in various visual recognition tasks, thereby supporting its adoption as a backbone model in studies requiring robust pattern recognition (Yu et al., 2024).

Due to its robust design, efficient training capability, and strong generalization, ResNet50 has frequently been chosen as a backbone network in computer vision studies requiring accurate pattern recognition under challenging conditions.

MobileNetV2

MobileNetV2 is a lightweight convolutional neural network architecture designed for efficient image classification and feature extraction on mobile and embedded devices. Introduced by Google in 2018, it improves upon MobileNetV1 by incorporating two key innovations: inverted residual blocks and linear bottlenecks (Altal et al., 2025; Li et al., 2024).

Architecture uses depthwise separable convolutions to reduce the number of parameters and computational complexity. In this operation, a standard convolution is decomposed into two sequential steps: a depthwise convolution applied with a single filter per input channel, followed by a pointwise (1×1) convolution that aggregates the resulting feature maps (Altal et al., 2025; Peng et al., 2024). These modifications significantly reduce the floating-point operations required, while preserving classification accuracy (Al-Gaashani et al., 2025).

Another crucial component of MobileNetV2 is the inverted residual structure, which differs from conventional residual blocks in that it first expands the feature dimensions before compressing them. This design helps retain feature information that may otherwise be lost in low-dimensional transformations. Additionally, linear bottlenecks are used in place of non-linear activations in the final layer of these blocks to minimize the loss of low-dimensional feature information (Li et al., 2024).

Shortcut connections between bottlenecks facilitate efficient gradient flow and feature reuse, improving both training dynamics and overall model performance (Al-Gaashani et al., 2025; Rokhva et al., 2024). As a result, MobileNetV2 achieves a favorable balance between speed, scalability, and generalization capability without significant compromise in classification accuracy, making it highly suitable for real-time surveillance and other resource-constrained environments (Zhou et al., 2024; Zou et al., 2025).

EfficientNet-B0

EfficientNet-B0 serves as the baseline model in the EfficientNet family, designed to provide an optimal trade-off between accuracy and computational efficiency. It was developed by Tan and Le through Neural Architecture Search (NAS), which enabled the automatic discovery of network depth, width, resolution, and architectural blocks under a fixed resource budget (Li et al., 2023; Xu et al., 2023).

The core architectural component of EfficientNet-B0 is the Mobile Inverted Bottleneck Convolution (MBConv) block, which is stacked in different configurations across seven main stages. Each MBConv block integrates depthwise separable convolutions, inverted residual connections, and Squeeze-and-Excitation (SE) modules, enabling efficient feature transformation while preserving representational power (Muthulakshmi et al., 2025; Rautaray et al., 2025; Wang et al., 2025).

Unlike conventional residual blocks, MBConv first expands the number of channels using pointwise convolutions, applies depthwise convolutions, and then projects the output back to a lower-dimensional space. This inverted residual structure improves efficiency while maintaining the ability to extract complex features (Canayaz, 2021; Li et al., 2023).

The compound scaling method employed in EfficientNet-B0 enables simultaneous, balanced scaling of depth, width, and resolution. This contrasts with traditional approaches that scale these dimensions independently. By quantifying their interdependencies through empirical analysis, compound scaling ensures optimal performance under resource constraints (Argho et al., 2024; P. Gangan et al., 2022; Yao et al., 2024).

The architecture begins with a 3×3 convolutional layer for initial feature extraction, followed by a sequence of MBConv blocks with 3×3 or 5×5 kernels, and concludes with a pointwise convolution, global average pooling, and a fully connected layer. The Swish activation function is used throughout the network, contributing to improved training dynamics and gradient flow (Rautaray et al., 2025; Yao et al., 2024).

EfficientNet-B0 has been validated across multiple benchmark datasets, including ImageNet and CIFAR-100, showing strong generalization performance. Its lightweight and modular design makes it particularly well-suited for deployment in resource-limited environments such as mobile or embedded devices (Aminanto et al., 2022; Himel et al., 2024; Wang et al., 2025).

Model Training Procedure

All models were implemented using the PyTorch framework and executed within a Kaggle Notebook environment. Each CNN was fine-tuned on the training dataset for the behavior recognition task. Transfer learning was employed: the convolutional base of each network, pre-trained on ImageNet, was retained, while the top classification layer was replaced with a newly defined fully connected layer containing 10 output units.

During training, the Adam optimizer (adaptive moment estimation) was used with a learning rate set to 0.0001 across all models. A mini-batch size of 32 was adopted. The categorical cross-entropy loss function served as the optimization criterion.

Rather than fixing the number of training epochs in advance, an early stopping strategy was employed to determine the optimal stopping point for each model. Training was allowed to proceed for a maximum of 50 epochs, with a patience parameter set to 5. If the validation loss failed to improve for five consecutive epochs, training was automatically terminated, and the model weights corresponding to the lowest validation loss were restored for subsequent evaluation. This approach eliminates the need for arbitrary epoch selection and

provides empirical evidence of convergence, as the stopping point is determined by the model's own learning dynamics rather than a predefined threshold.

Evaluation Metrics

To evaluate model performance in the multi-class behavior classification task, confusion matrices and derived classification metrics were utilized. A confusion matrix summarizes the model's predictions by counting True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), providing a foundation for metric computation.

From these quantities, key evaluation metrics were calculated: accuracy, precision, recall, and F1-score.

- Accuracy represents the overall proportion of correct predictions and is computed as the ratio of correctly classified instances to the total number of predictions. While it provides a general performance overview, its utility decreases in the presence of class imbalance.
- Precision (Positive Predictive Value) indicates how many of the predicted positive cases are actually correct. In the context of surveillance systems, high precision is crucial to minimize false alarms and avoid misclassifying benign behaviors as threats.
- Recall (Sensitivity or True Positive Rate) measures the model's ability to detect all instances of a specific class. For safety-critical applications like violence detection, high recall ensures that potentially dangerous behaviors are less likely to be missed.
- F1-score, defined as the harmonic mean of precision and recall, balances these two metrics and is particularly useful when one seeks to minimize both false positives and false negatives.

The corresponding formulas are presented in Equations (1)–(4) (AbuAlkebash et al., 2025):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4}$$

Each metric was computed individually for all ten behavior classes, and macro-averaged values were reported to capture overall model performance across categories. Additionally, weighted averages were calculated to reflect the influence of class distribution on the overall evaluation.

To provide a more detailed performance analysis, confusion matrices were visualized as heatmaps, highlighting misclassification patterns across behavior classes. Furthermore, ROC (Receiver Operating Characteristic) curves were employed to evaluate discriminative capability. For each class, a one-vs-rest strategy was adopted, where the given class was treated as positive and the others as negative, allowing individual ROC curves and Area Under the Curve (AUC) values to be computed.

A combined ROC plot was generated for each CNN model, and the macro-average AUC was reported as an aggregate measure of the classifier's ability to distinguish among classes. High AUC values (close to 1.0) indicate robust class separability. The AUC-ROC metric is particularly valuable in class-imbalanced contexts, as it remains unaffected by class frequency and enables more reliable model comparison (Redondo et al., 2026). In practice, this visualization also supports application-specific threshold tuning, for example, prioritizing high recall in detecting violent behavior, even at the cost of increased false positives.

Results

The training process for all three CNN architectures, EfficientNet-B0, ResNet50, and MobileNetV2, was monitored using an early stopping strategy with a patience of 5, with the best model weights restored based on the lowest validation loss achieved during training. Both training and validation loss values were recorded at each epoch. As presented in Table 1, each model exhibited a sharp decrease in training loss within the first few epochs, after which the reduction gradually stabilized, indicating convergence toward a minimum. Validation loss followed a similar downward trend, confirming that the models generalized well to unseen data. This trend is further illustrated in Figure 1, where both loss curves visibly flatten as training progresses.

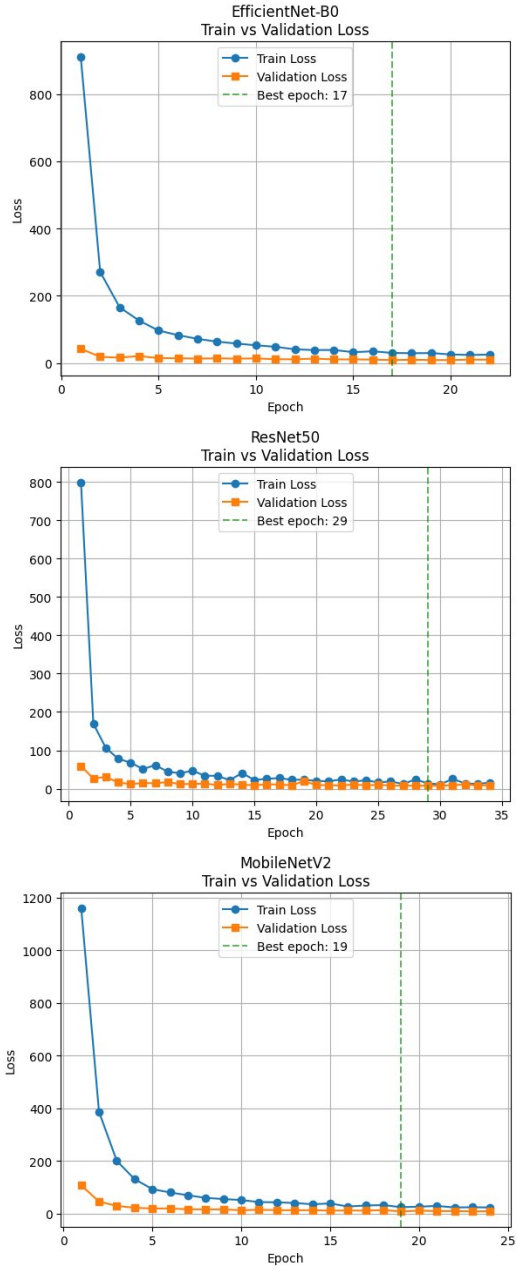

Figure 1. Training and validation losses per epoch for EfficientNet-B0, ResNet50, and MobileNetV2 models

Table 1. Epoch-based training and validation losses for different CNN architectures

Architecture	Best Epoch No	Training Loss	Validation Loss
EfficientNet-B0	17	29.8799	8.4837
ResNet50	29	13.6272	7.5118
MobileNetV2	19	23.8092	7.5944

MobileNetV2 began with the highest initial training loss (1160.43), followed by EfficientNet-B0 (911.16) and ResNet50 (798.84). Despite this, all three models demonstrated rapid convergence in the early epochs. ResNet50 required the most epochs to reach its best checkpoint (epoch 29) but ultimately achieved the lowest validation loss (7.51), suggesting that its deeper architecture benefited from extended training under the early stopping regime. EfficientNet-B0 converged earliest among the three at epoch 17 with a best validation loss of 8.48, while MobileNetV2 reached its best performance at epoch 19 with a validation loss of 7.59. These results confirm that transfer learning enabled all models to converge effectively, and the consistently low validation losses across architectures indicate that overfitting was successfully mitigated throughout training.

All models were subsequently evaluated on a test set comprising 6,711 video instances. The evaluation was conducted using widely adopted classification metrics, including accuracy, precision, recall, F1-score, and AUC. Performance metrics for each model are reported class-wise, along with macro-averaged and weighted-averaged results to ensure balanced comparisons. Additionally, confusion matrices and ROC curves were employed to provide visual insights into each model's classification behavior.

Table 2 presents the classification results of EfficientNet-B0 on the test dataset. The model achieved high precision and recall across all behavior classes. Notably, Normal, Shoplifting, Fighting, and Stealing each achieved precision and recall values of 0.99 or higher, reflecting the strongest per-class performance. The lowest precision was observed in Burglary (0.96), while the lowest recall values were recorded for Shooting (0.96), though these still reflect strong performance. Overall, the model achieved 99% accuracy and a macro-average F1 Score of 0.98, indicating robust, balanced classification performance across all 10 behavior categories.

Table 2. Classification performance metrics of EfficientNet-B0 model on multi-class surveillance dataset

Class	Precision	Recall	F1-Score	Support
Abuse	0.98	0.98	0.98	583
Assault	1.00	0.97	0.98	383
Burglary	0.96	1.00	0.98	561
Fighting	0.99	0.99	0.99	788
Normal	1.00	0.99	0.99	1377
Robbery	0.99	0.97	0.98	405
Shooting	0.98	0.96	0.97	447
Shoplifting	0.99	1.00	1.00	960
Stealing	0.99	0.99	0.99	760
Vandalism	0.99	0.98	0.98	448
Accuracy			0.99	6712
Macro avg	0.99	0.98	0.98	6712
Weighted avg	0.99	0.99	0.99	6712

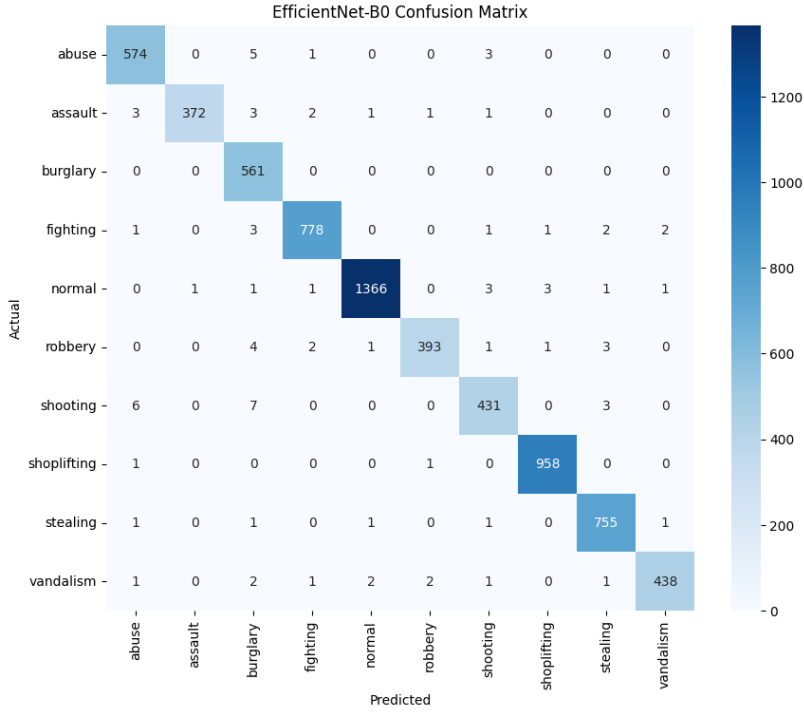

Figure 2. Confusion matrix of the EfficientNet-B0 model on the multi-class surveillance dataset

As illustrated in Figure 2, the confusion matrix demonstrates that the EfficientNet-B0 model accurately classified the majority of instances across all 10 behavior categories. The diagonal dominance indicates strong agreement between predicted and true labels, with minimal misclassifications observed. Notably, Burglary achieved perfect classification with no misclassified instances, while Shoplifting also demonstrated near-perfect performance with only 2 errors. The most notable misclassifications occurred in Shooting, where several instances were incorrectly assigned to the Abuse and Burglary classes, and in Vandalism and Assault, which exhibited slightly higher confusion across multiple categories. Overall, these results confirm the model's robustness and reliability in distinguishing between diverse behavior classes in real-world surveillance scenarios.

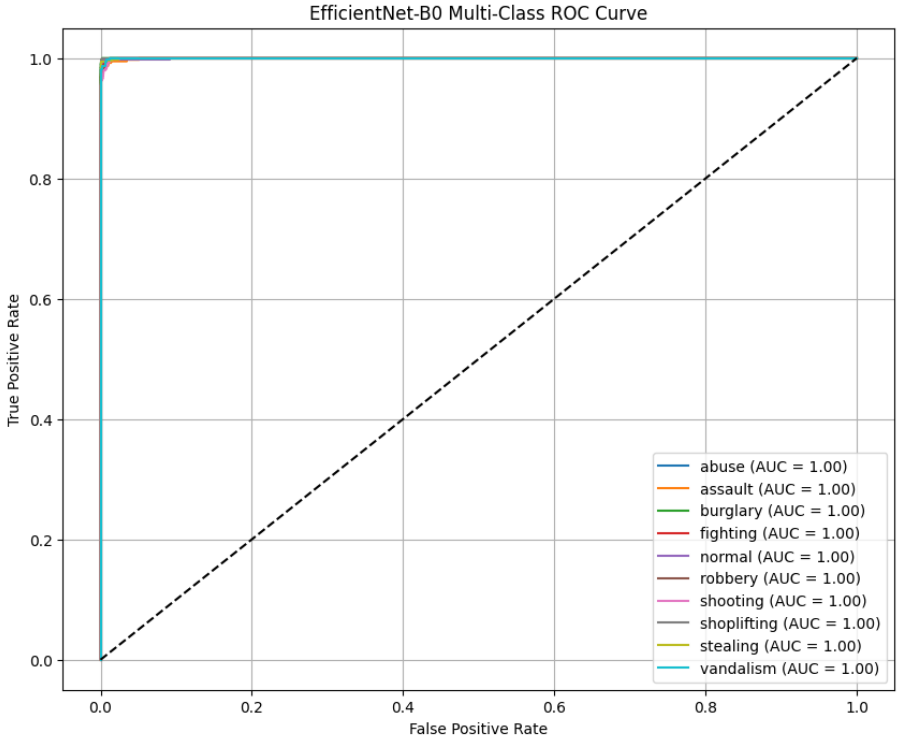

Figure 3. Multi-class ROC curve of the EfficientNet-B0 model on the surveillance dataset

Figure 3 presents the multi-class ROC curves of the EfficientNet-B0 model, generated using a one-vs-rest strategy. All ten behavior classes achieved a perfect AUC score of 1.00, with each curve reaching the top-left corner of the plot, indicating near-zero false-positive rates alongside maximum true-positive rates. This result demonstrates exceptional discriminative ability across all categories,

confirming that the model can reliably distinguish each behavior class from the rest. The uniformity of AUC scores across all classes further highlights the model's consistent and balanced classification performance, regardless of class-specific characteristics.

Table 3 presents the classification performance of the ResNet50 model on the test dataset. The model achieved exceptionally high precision and recall across all behavior classes. Notably, Shoplifting and Burglary both achieved a perfect F1-score of 1.00, while classes such as Abuse, Fighting, Normal, Robbery, Shooting, and Stealing each reached an F1-score of 0.99. The lowest F1-score was observed in Vandalism and Assault (0.98), which still reflects strong performance. Overall, ResNet50 achieved 99% accuracy and a macro-average F1-score of 0.99, demonstrating outstanding and balanced classification performance across all 10 behavior categories.

Table 3. Classification performance metrics of ResNet50 model on multi-class surveillance dataset

Class	Precision	Recall	F1-Score	Support
Abuse	1.00	0.98	0.99	583
Assault	0.98	0.98	0.98	383
Burglary	1.00	0.99	1.00	561
Fighting	0.98	0.99	0.99	788
Normal	0.99	0.99	0.99	1377
Robbery	0.99	0.99	0.99	405
Shooting	1.00	0.98	0.99	447
Shoplifting	1.00	1.00	1.00	960
Stealing	0.99	1.00	0.99	760
Vandalism	0.99	0.98	0.98	448
Accuracy			0.99	6712
Macro avg	0.99	0.99	0.99	6712
Weighted avg	0.99	0.99	0.99	6712

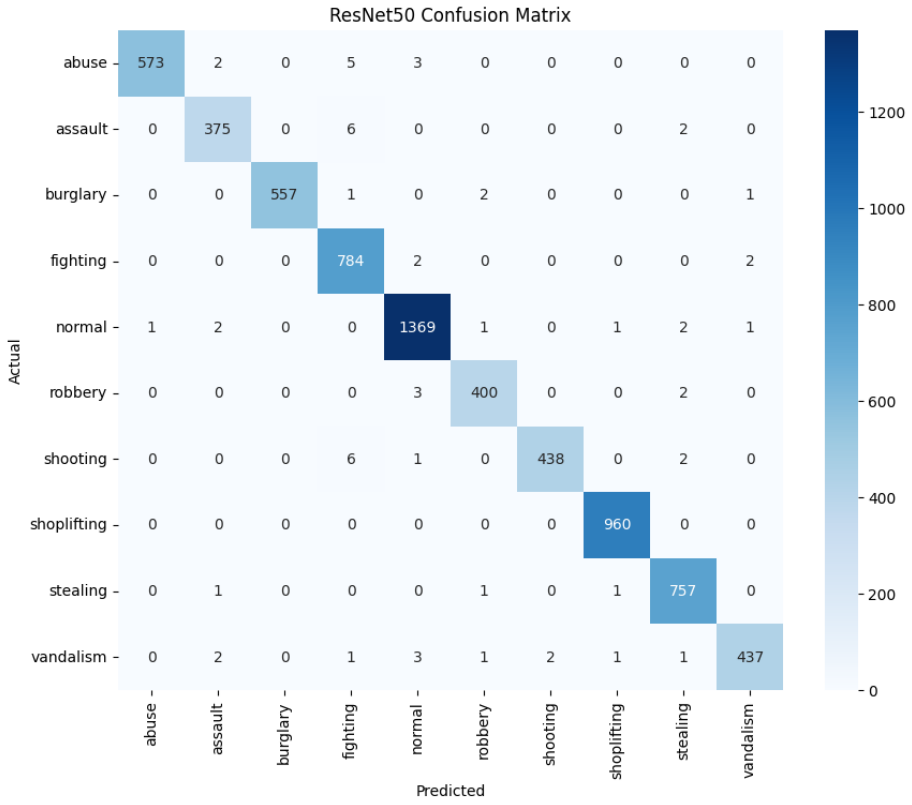

Figure 4. Confusion matrix of the ResNet50 model on the multi-class surveillance dataset

As illustrated in Figure 4, the ResNet50 model accurately classified the majority of test samples across all ten behavior classes, as evidenced by the strong diagonal alignment. Shoplifting achieved perfect classification with no misclassifications, while Normal and Burglary also showed minimal confusion. The most notable misclassifications were observed in Vandalism (11 errors distributed across multiple classes) and Shooting (9 errors, predominantly misclassified as Fighting), indicating these categories present greater classification challenges. Overall, the results confirm the model's strong and reliable performance across diverse surveillance behavior classes.

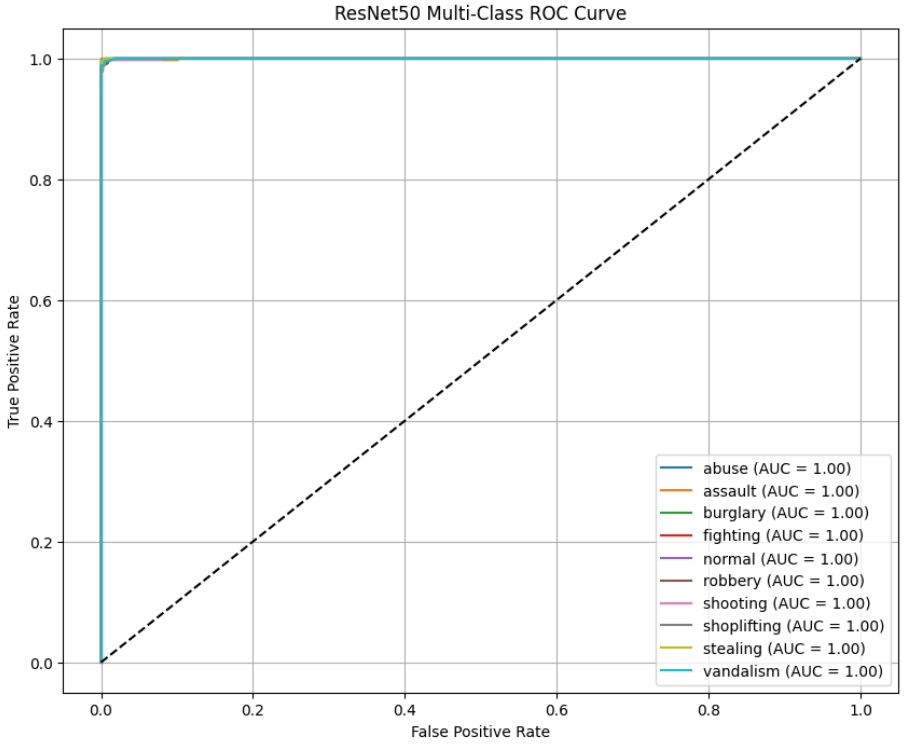

Figure 5. Multi-class ROC curve of the ResNet50 model on the surveillance dataset

The ROC curves in Figure 5 demonstrate the exceptional discriminative capability of the ResNet50 model across all ten behavior classes. The AUC for each class reached 1.00, indicating perfect separation between positive and negative instances in the one-vs-rest classification setup. These results further validate the model's outstanding classification performance, consistent with the high precision, recall, and F1-scores reported in Table 3.

Table 4 presents the classification results obtained from the MobileNetV2 model. The model achieved 99% overall accuracy and a macro-average F1-score of 0.99, demonstrating strong and consistent performance across all behavior categories. Shoplifting achieved a perfect F1-score of 1.00, while classes such as Abuse, Burglary, Normal, and Stealing each reached 0.99. The lowest F1-score was observed in Shooting (0.97), though this still reflects high classification performance. Overall, MobileNetV2 demonstrated competitive results comparable to EfficientNet-B0 and ResNet50, despite being a more lightweight architecture.

Table 4. Classification performance metrics of MobileNetV2 model on multi-class surveillance dataset

Class	Precision	Recall	F1-Score	Support
Abuse	0.99	0.98	0.99	583
Assault	0.98	0.97	0.98	383
Burglary	0.99	0.99	0.99	561
Fighting	0.97	0.99	0.98	788
Normal	0.99	0.99	0.99	1377
Robbery	0.98	0.99	0.98	405
Shooting	0.97	0.97	0.97	447
Shoplifting	1.00	1.00	1.00	960
Stealing	0.99	1.00	0.99	760
Vandalism	0.99	0.97	0.98	448
Accuracy			0.99	6712
Macro avg	0.99	0.98	0.99	6712
Weighted avg	0.99	0.99	0.99	6712

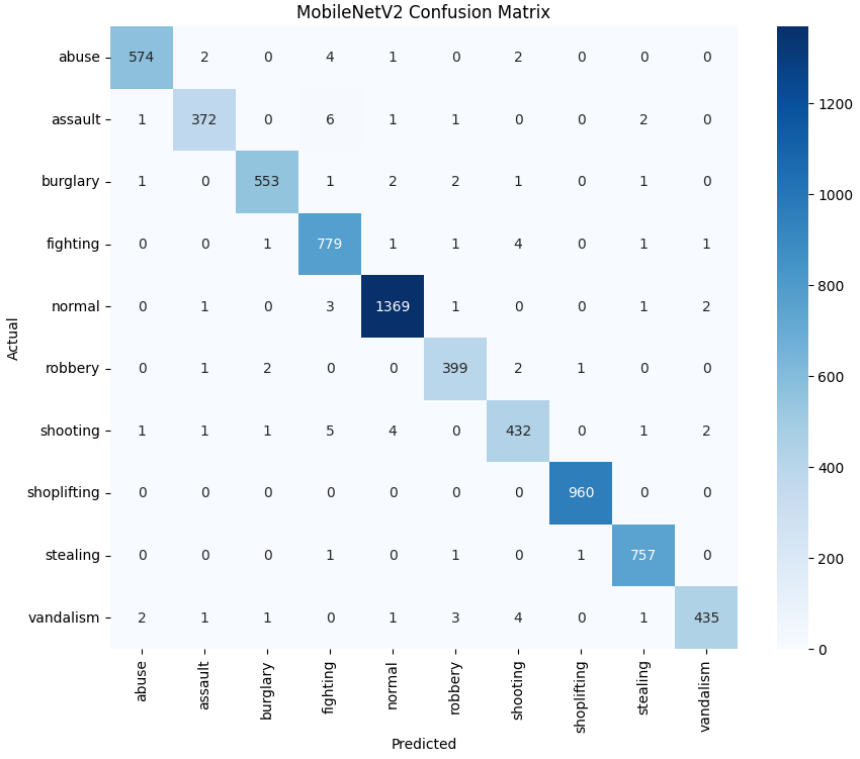

Figure 6. Confusion matrix of the MobileNetV2 model on the multi-class surveillance dataset

Figure 6 illustrates the confusion matrix for the MobileNetV2 model. A clear diagonal trend indicates strong overall performance, with the majority of samples correctly classified across all ten behavior categories. Shoplifting achieved perfect classification with no misclassifications, while Normal and Stealing also showed minimal confusion. The most notable misclassifications were observed in Shooting (15 errors, predominantly misclassified as Fighting) and Vandalism (13 errors distributed across multiple classes), suggesting these categories present greater classification challenges for the model. Nevertheless, MobileNetV2 demonstrated reliable and efficient performance across most behavior classes.

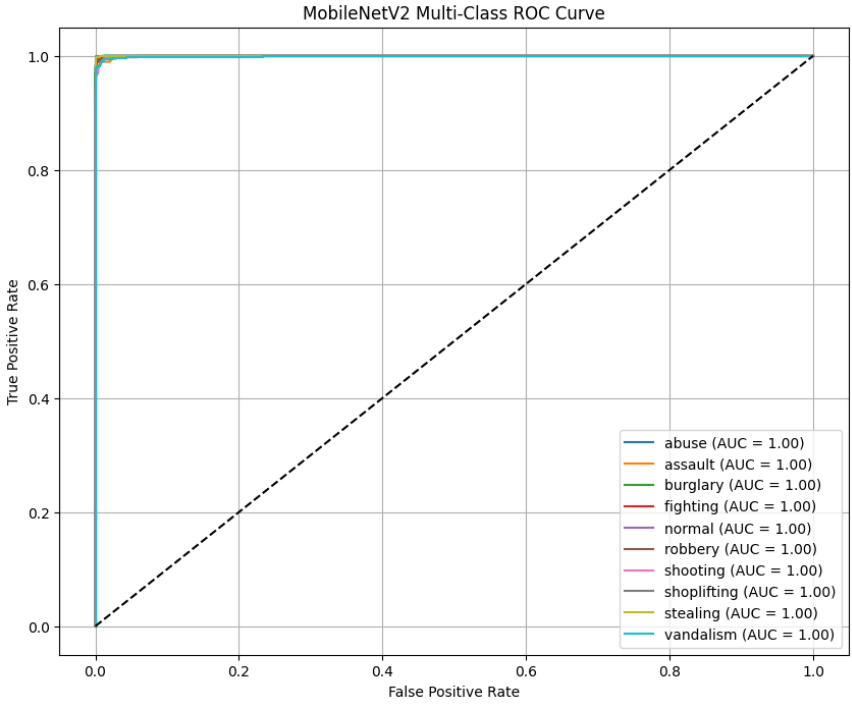

Figure 7. Multi-class ROC curve of the MobileNetV2 model on the surveillance dataset

The ROC curves in Figure 7 further confirm the exceptional discriminative capability of the MobileNetV2 model. All ten behavior classes achieved a perfect AUC score of 1.00, indicating complete separation between positive and negative instances in the one-vs-rest classification setup. These results validate the model's outstanding classification performance, consistent with the high precision, recall, and F1-scores reported in Table 4, and further highlight its potential as a lightweight yet powerful architecture for real-time surveillance applications.

Table 5 presents the overall classification performance of three CNN architectures, ResNet50, EfficientNet-B0, and MobileNetV2, across five key metrics. ResNet50 achieved the best results across all metrics, with the highest accuracy (0.9908), macro precision (0.9907), macro recall (0.9883), macro F1-score (0.9895), and macro AUC (0.9999), demonstrating superior and consistent performance. MobileNetV2 ranked second with an accuracy of 0.9878 and a macro F1-score of 0.9851, offering competitive performance despite its lightweight architecture. EfficientNet-B0 followed closely with an accuracy of 0.9872 and a macro F1-score of 0.9847. Overall, all three models achieved high, comparable results, confirming their effectiveness in surveillance-based behavior classification.

Table 5. Comparison of overall performance metrics across CNN architectures

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-score	Macro AUC
ResNet50	**0.9908**	**0.9907**	**0.9883**	**0.9895**	**0.9999**
MobileNetV2	0.9878	0.9860	0.9843	0.9851	0.9998
EfficientNet-B0	0.9872	0.9856	0.9839	0.9847	0.9998

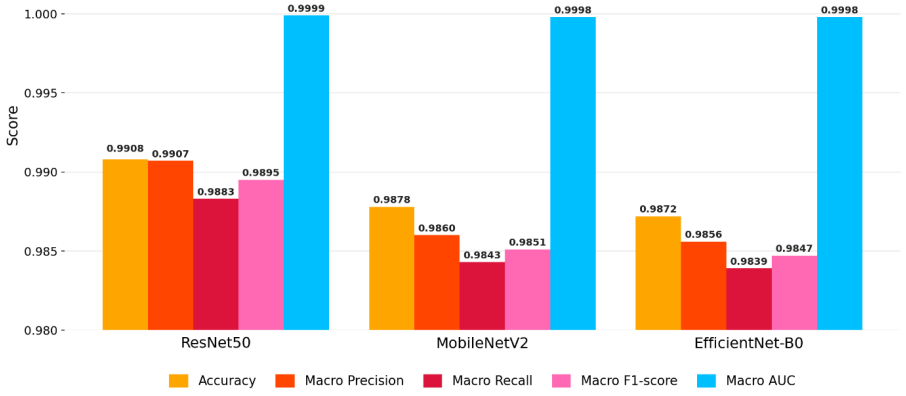

Figure 8. Bar chart of classification performance metrics for CNN architectures

Figure 8 offers a visual comparison of the five-evaluation metrics across the three CNN models. As clearly illustrated, ResNet50 consistently outperformed the other architectures across all metrics, followed by MobileNetV2 and EfficientNet-B0 in close succession. Notably, all three models achieved comparable, high scores, with differences between them remaining marginal. The bar chart further highlights that macro AUC values were uniformly near-perfect across all models, reinforcing their strong discriminative capability for multi-class surveillance behavior classification.

Discussion

This study compared three CNN architectures, EfficientNet-B0, ResNet50, and MobileNetV2, for multi-class behavior recognition in surveillance videos. Among them, ResNet50 demonstrated the best overall performance, achieving the highest accuracy, macro F1-score, and macro AUC. Its deep residual architecture likely contributed to effective feature extraction, enabling strong generalization across all behavior categories. MobileNetV2 ranked second across all metrics, and despite its lightweight architecture and reduced computational demands, delivered highly competitive results, making it a strong candidate for resource-constrained deployment scenarios such as smart surveillance cameras or embedded systems. EfficientNet-B0 followed closely in third place, with its compound scaling mechanism enabling efficient parameter utilization and maintaining highly competitive performance for real-world surveillance applications.

An analysis of misclassifications revealed that confusion often occurred between semantically similar classes, e.g., Shooting and Fighting, or Vandalism and other action-based behaviors. These overlaps are not unexpected and can even challenge human judgment without contextual cues. Such misclassifications were consistently observed across all three models, suggesting they reflect inherent ambiguity in the visual appearance of certain behavior categories rather than model-specific limitations.

From an operational standpoint, the choice of model depends on deployment constraints. ResNet50 offers the highest predictive performance but may require greater computational resources. MobileNetV2 provides a viable alternative with only marginal accuracy trade-offs but significantly reduced computational demands. These findings highlight the importance of selecting architectures based on both predictive performance and infrastructure feasibility in real-world surveillance systems.

Conclusion

This study presented a comprehensive evaluation of three convolutional neural network (CNN) architectures, EfficientNet-B0, ResNet50, and MobileNetV2, for the task of automated behavior recognition in private security surveillance. ResNet50 consistently outperformed the others across all key metrics, achieving the highest accuracy (99.08%), macro F1-score (0.9895), and macro AUC (0.9999). These results highlight the effectiveness of deep residual architectures in extracting discriminative features from surveillance footage. MobileNetV2 and EfficientNet-B0 also demonstrated strong classification capabilities, with accuracies of 98.78% and 98.72%, respectively, confirming that all three architectures are highly effective for this task.

While ResNet50 achieved the highest overall performance, MobileNetV2 offers a compelling trade-off between performance and computational efficiency. Its lightweight structure and reduced parameter count make it a practical choice for edge-based or resource-constrained deployments, where real-time analysis on multiple camera streams is needed. These findings confirm the potential of CNN-based AI systems to act as force multipliers in surveillance operations, enabling faster response times and reducing human workload.

Despite promising results, some limitations must be acknowledged. First, the models were trained and evaluated exclusively on a single dataset (UCF-Crime). Cross-dataset evaluation using a second benchmark dataset would have provided stronger evidence of generalizability. Future studies should incorporate cross-dataset testing protocols to assess whether the learned representations transfer effectively across different surveillance environments, camera configurations, and behavior distributions. The evaluation was conducted on a fixed dataset that may not fully capture the diversity of real-world environments. Variations in lighting, camera placement, resolution, and behavior types may affect generalization performance if not properly addressed. Furthermore, the current approach operates on individual frames or short segments, without incorporating temporal modeling. As many human behaviors unfold over time, the lack of temporal context could lead to confusion between similar actions, such as shooting and fighting. Additionally, multi-label situations, where multiple behaviors occur simultaneously, were not considered in the current classification setup.

Future research should explore integrating temporal architectures, such as 3D CNNs, LSTMs, or video transformers, to better capture motion dynamics. Domain adaptation techniques may also be employed to enhance robustness across diverse deployment contexts. Real-world pilot implementations in varied environments such as retail stores, campuses, or public transport hubs could provide valuable insights into operational challenges and help calibrate model sensitivity. Furthermore, explainability tools like Grad-CAM can be integrated to enhance transparency and user trust by highlighting the regions contributing most to predictions. Finally, expanding the behavior taxonomy and supporting multi-label classification will enable models to better reflect the complexity of real-world surveillance footage.

In summary, ResNet50 emerges as the most accurate solution for multi-class behavior recognition in surveillance videos, while MobileNetV2 stands out as the most practical option for resource-constrained deployments. The deployment of these architectures in AI-powered surveillance systems could significantly improve incident detection, response time, and overall situational awareness in private security settings. Continued research and refinement will be essential to further advance these capabilities and ensure reliable performance in diverse and dynamic environments.

Declaration of Competing interests

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The dataset employed in this study is publicly accessible and may be used freely for research and academic purposes. It can be obtained from Real-Time Anomaly Detection in CCTV Surveillance on Kaggle, available at https://www.kaggle.com/datasets/webadvisor/real-time-anomaly-detection-in-cctv-surveillance.

References

AbuAlkebash, H., Saleh, R. A. A., & Ertunç, H. M. (2025). Automated explainable deep learning framework for multiclass skin cancer detection and classification using hybrid YOLOv8 and vision transformer (ViT). *Biomedical Signal Processing and Control*, *108*, 107934. https://doi.org/10.1016/j.bspc.2025.107934

Ahmed, I., & Naib, B. B. (2023, 29–30 April 2023). Object Motion Tracking and Detection in Surveillance Videos using Resnet Architecture. 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE),

Akshith, N., Adithya, S. K., Abhiram, M., Vikas, M., Achuth, D., & Murthy, G. N. (2022, 22–23 Dec. 2022). Action Recognition with Neural Networks. 2022 Sardar Patel International Conference on Industry 4.0 - Nascent Technologies and Sustainability for 'Make in India' Initiative,

Al-Gaashani, M. S. A. M., Xu, W., & Obsie, E. Y. (2025). MobileNetV2-based deep learning architecture with progressive transfer learning for accurate monkeypox detection. *Applied Soft Computing*, *169*, 112553. https://doi.org/10.1016/j.asoc.2024.112553

Alamuru, S., & Jain, S. (2022, 29–30 July 2022). A Quick Review and Performance Analysis of Custom and Transfer Learning CNN Architectures for Event Detection in Videos. 2022 IEEE International Conference on Data Science and Information System (ICDSIS),

Altal, O. F., Sindiani, A. M., Amin, M., Abu Mhanna, H. Y., Hamad, R., Gharaibeh, H., Akhdar, H. F., Alhatamleh, S., Almahmoud, R. E., Abu-azzam, O. H., Balaw, M., Hamoud, B. H., Maashey, F., & Alghulayqah, L. (2025). Hybrid attention-enhanced MobileNetV2 with particle swarm optimization for endometrial cancer classification in CT images. *Informatics in Medicine Unlocked*, *57*, 101662. https://doi.org/10.1016/j.imu.2025.101662

Amin, S. U., Abbas, M. S., Kim, B., Jung, Y., & Seo, S. (2024). Enhanced Anomaly Detection in Pandemic Surveillance Videos: An Attention Approach With EfficientNet-B0 and CBAM Integration. *IEEE Access*, *12*, 162697–162712. https://doi.org/10.1109/ACCESS.2024.3488797

Aminanto, M. E., Purbomukti, I. R., Chandra, H., & Kim, K. (2022). Two-Dimensional Projection-Based Wireless Intrusion Classification Using Lightweight EfficientNet. *Computers, Materials and Continua*, *72*(3), 5301–5314. https://doi.org/10.32604/cmc.2022.026749

Ansari, M. A., & Singh, D. K. (2022). An Expert Eye for Identifying Shoplifters in Mega Stores. In A. Khanna, D. Gupta, S. Bhattacharyya, A. E. Hassanien, S. Anand, & A. Jaiswal, *International Conference on Innovative Computing and Communications* Singapore.

Argho, A. G., Maswood, M. M. S., Mahmood, M. I., & Mondol, N. (2024). Efficient-CovNet: A CNN-based approach to detect various pulmonary diseases including COVID-19 using modified EfficientNet. *Intelligent Systems with Applications*, *21*, 200315. https://doi.org/10.1016/j.iswa.2023.200315

Bhimavarapu, J. P., Ramaraju, S., Nagajyothi, D., & Rao, I. V. (2023). Convolutional neural network based object detection system for video surveillance application. *Concurrency and Computation: Practice and Experience*, *35*(3), e7461. https://doi.org/10.1002/cpe.7461

Bohlol, P., Hosseinpour, S., & Soltani Firouz, M. (2025). Improved food recognition using a refined ResNet50 architecture with improved fully connected layers. *Current Research in Food Science*, *10*, 101005. https://doi.org/10.1016/j.crfs.2025.101005

Canayaz, M. (2021). C+EffxNet: A novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet. *Chaos, Solitons & Fractals*, *151*, 111310. https://doi.org/10.1016/j.chaos.2021.111310

Duong, H. T., Le, V. T., & Hoang, V. T. (2023). Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. *Sensors (Basel)*, *23*(11). https://doi.org/10.3390/s23115024

Elmetwally, A., Eldeeb, R., & Elmougy, S. (2025). Deep learning based anomaly detection in real-time video. *Multimedia Tools and Applications*, *84*(11), 9555–9571. https://doi.org/10.1007/s11042-024-19116-9

Himel, G. M. S., Islam, M. M., & Rahaman, M. (2024). Utilizing EfficientNet for sheep breed identification in low-resolution images. *Systems and Soft Computing*, *6*, 200093. https://doi.org/10.1016/j.sasc.2024.200093

Jaggi, A., Aggarwal, A., & Gupta, A. (2023, 3–4 March 2023). Identifying Anti-Social Activities in Surveillance Monitoring Applications using Deep-CNN based Algorithms. 2023 6th International Conference on Information Systems and Computer Networks (ISCON),

Kaggle. (2025). *Real Time Anomaly Detection in CCTV Surveillance*. https://www.kaggle.com/datasets/webadvisor/real-time-anomaly-detection-in-cctv-surveillance

Kerachi, S., Koma, A. K., & Asharioun, H. (2023, 25–26 Jan. 2023). Real-time Steal Recognition on CCTV-Based Videos for Embedded Systems. 2023 28th International Computer Conference, Computer Society of Iran (CSICC),

Khanam, S., Sharif, M., Cheng, X., & Kadry, S. (2024). Suspicious action recognition in surveillance based on handcrafted and deep learning methods: A survey of the state of the art. *Computers and Electrical Engineering*, *120*, 109811. https://doi.org/10.1016/j.compeleceng.2024.109811

Li, J., Zeng, H., Huang, C., Wu, L., Ma, J., Zhou, B., Ye, D., & Weng, H. (2023). Noninvasive Detection of Salt Stress in Cotton Seedlings by Combining Multicolor

Fluorescence–Multispectral Reflectance Imaging with EfficientNet-OB2. *Plant Phenomics*, *5*, 0125. https://doi.org/10.34133/plantphenomics.0125

Li, Z., Li, Y., Yan, C., Yan, P., Li, X., Yu, M., Wen, T., & Xie, B. (2024). Enhancing Tea Leaf Disease Identification with Lightweight MobileNetV2. *Computers, Materials and Continua*, *80*(1), 679–694. https://doi.org/10.32604/cmc.2024.051526

Liao, C., Pu, C., Chen, T., Todo, Y., Furuichi, K., Yabe, T., & Qiu, D. (2025). Image analysis of diabetes pathology: classifying with precision via an upgraded resnet50 framework. *Medicine in Novel Technology and Devices*, *27*, 100385. https://doi.org/10.1016/j.medntd.2025.100385

Lu, Y., Xu, Z., & Wang, J. (2022, 24–26 June 2022). Abnormal Behavior Recognition System based on Improved CRNN Model. 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA),

Mahjoubi, M. A., Lamrani, D., Saleh, S., Moutaouakil, W., Ouhmida, A., Hamida, S., Cherradi, B., & Raihani, A. (2025). Optimizing ResNet50 performance using stochastic gradient descent on MRI images for Alzheimer's disease classification. *Intelligence-Based Medicine*, *11*, 100219. https://doi.org/10.1016/j.ibmed.2025.100219

Mohanaprakash, T. A., Somu, C. S., Nirmalrani, V., Vyshnavi, K., Sasikumar, A. N., & Shanthi, P. (2024, 18–20 Sept. 2024). Detection of Abnormal Human Behavior using YOLO and CNN for Enhanced Surveillance. 2024 5th International Conference on Smart Electronics and Communication (ICOSEC),

Muthulakshmi, M., Venkatesan, K., Prasanalakshmi, B., Syarifah Bahiyah, R., & Divya, V. (2025). An intelligent ensemble EfficientNet prediction system for interpretations of cardiac magnetic resonance images in heart failure severity diagnosis. *Intelligence-Based Medicine*, *11*, 100218. https://doi.org/10.1016/j.ibmed.2025.100218

Natha, S., Ahmed, F., Siraj, M., Lagari, M., Altamimi, M., & Chandio, A. A. (2025). Deep BiLSTM Attention Model for Spatial and Temporal Anomaly Detection in Video Surveillance. *Sensors*, *25*(1), 251. https://doi.org/10.3390/s25010251

Ngoc, H. N., Xuan, N. N., Bui, T. H., Hung, D. H., Truong, S. Q. H., & Hoang, V. (2023, 5–8 Jan. 2023). An efficient approach for real-time abnormal human behavior recognition on surveillance cameras. 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG),

P. Gangan, M., K, A., & V. L, L. (2022). Distinguishing natural and computer generated images using Multi-Colorspace fused EfficientNet. *Journal of Information Security and Applications*, *68*, 103261. https://doi.org/10.1016/j.jisa.2022.103261

Pathirannahalage, I., Jayasooriya, V., Samarabandu, J., & Subasinghe, A. (2025). A comprehensive analysis of real-time video anomaly detection methods for human and vehicular movement. *Multimedia Tools and Applications*, *84*(10), 7519–7564. https://doi.org/10.1007/s11042-024-19204-w

Peng, L., Zhang, J., Li, Y., & Du, G. (2024). A novel percussion-based approach for pipeline leakage detection with improved MobileNetV2. *Engineering Applications of Artificial Intelligence, 133*, 108537. https://doi.org/10.1016/j.engappai.2024.108537

Rath, A., Mishra, B. S. P., & Bagal, D. K. (2025). ResNet50-based Deep Learning model for accurate brain tumor detection in MRI scans. *Next Research*, *2*(1), 100104. https://doi.org/10.1016/j.nexres.2024.100104

Rautaray, J., Ali, A. B. M., Kandpal, M., Mishra, P., Rashid, R. F., Alimova, F., Kallel, M., & Batool, N. (2025). Leveraging FastViT based knowledge distillation with EfficientNet-B0 for diabetic retinopathy severity classification. *SLAS Technology*, *33*, 100325. https://doi.org/10.1016/j.slast.2025.100325

Razak, H., Tahir, N. M., Abd Almisreb, A., Zakaria, N., & Zamri, N. (2022). Detection of criminal behavior at the residential unit based on deep convolutional neural network. *International Journal of Advanced Computer Science and Applications*, *13*(2).

Redondo, A., Ivaylova, K., Bachiller, M., Rincón, M., Cuadra, J. M., Tamimi, F., López-Cedrún, J. L., Diniz-Freitas, M., Lago-Méndez, L., Rubín-Roger, G., Torres, J., Bagán, L., Hernández, G., & López-Pintor, R. M. (2026). Multiclass classification of oral mucosal lesions by deep learning from clinical images without performing any restrictions. *Biomedical Signal Processing and Control*, *111*, 108337. https://doi.org/10.1016/j.bspc.2025.108337

Rokhva, S., Teimourpour, B., & Soltani, A. H. (2024). Computer vision in the food industry: Accurate, real-time, and automatic food recognition with pretrained MobileNetV2. *Food and Humanity*, *3*, 100378. https://doi.org/10.1016/j.foohum.2024.100378

Saket, O., Aicha, Anis B., & Fathallah, H. (2025). Deep learning applied for abnormal human behavior recognition in video surveillance systems: A systematic review. *Applied Intelligence*, *55*(13), 904. https://doi.org/10.1007/s10489-025-06797-4

Silva, D., Manzo-Martínez, A., Gaxiola, F., Gonzalez-Gurrola, L., & Ramírez-Alonso, G. (2022). Analysis of CNN architectures for human action recognition in video. *Computación y Sistemas*, *26*(2), 623–641.

Wang, Y., Sun, Y., Li, Y., & Zhou, C. (2025). Malicious Document Detection Based on GGE Visualization. *Computers, Materials and Continua*, *82*(1), 1233–1254. https://doi.org/10.32604/cmc.2024.057710

Xu, W., Nie, L., Chen, B., & Ding, W. (2023). Dual-stream EfficientNet with adversarial sample augmentation for COVID-19 computer aided diagnosis. *Computers in Biology and Medicine*, *165*, 107451. https://doi.org/10.1016/j.compbiomed.2023.107451

Yao, L., Liu, B., & Xin, Y. (2024). Visualization-based comprehensive feature representation with improved EfficientNet for malicious file and variant recognition. *Journal of Information Security and Applications*, *86*, 103865. https://doi.org/10.1016/j.jisa.2024.103865

Yu, Q., Zhang, Y., Xu, J., Zhao, Y., & Zhou, Y. (2024). Intelligent damage classification for tensile membrane structure based on continuous wavelet transform and improved ResNet50. *Measurement*, *227*, 114260. https://doi.org/10.1016/j.measurement.2024.114260

Zhou, G., He, Q., Liu, X., Kai, X., Cao, W., Ding, J., Zhuang, B., Xu, S., & Thwin, M. (2024). Optimizing MobileNetV2 for improved accuracy in early gastric cancer detection based on dynamic pelican optimizer. *Heliyon*, *10*(16), e35854. https://doi.org/10.1016/j.heliyon.2024.e35854

Zou, Y., Wu, L., Zuo, C., Chen, L., Zhou, B., & Zhang, H. (2025). White blood cell classification network using MobileNetv2 with multiscale feature extraction module and attention mechanism. *Biomedical Signal Processing and Control*, *99*, 106820. https://doi.org/10.1016/j.bspc.2024.106820