Journal of Rare Cardiovascular Diseases
ISSN: 2299-3711 (Print) | e-ISSN: 2300-5505 (Online)
www.jrcd.eu
RESEARCH ARTICLE

Real-Time Traffic Anomaly Detection Using Edge-Deployed LSTM Model on CCTV Surveillance Systems

P.M.Prinsar[1], Dr.A.Krishnaveni[2]
[1]Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore-641 021.
[2]Assistant Professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore-641 021.

*Corresponding Author

Article History
Received: 17.09.2025
Revised: 02.10.2025
Accepted: 15.10.2025
Published: 29.10.2025

Abstract: The ever-increasing number of vehicles leads to a drastic increase in real-time traffic, creating more opportunities and possibilities for anomalies in road traffic. Continuous monitoring solutions are necessary for anomaly detection, which can be achieved by developing and deploying an intelligent surveillance monitoring system that processes and predicts anomalies using efficient AI algorithms. This is particularly important for smart surveillance monitoring systems in road traffic applications. The surveillance cameras provide spatiotemporal data, which has two different complex dependencies that conventional learning algorithms cannot process. Thus, this paper proposes a Multi-CNN (Multiple Convolutional Neural Network) model to process and predict anomalies that occur in data generated by CCTV cameras deployed in the edge road network. Initially, a continuous video streaming acquisition task is applied to the CCTV cameras to create a dataset, which is then processed using Exploratory Data Analysis to understand the road traffic patterns. EDA identifies outliers and anomalies by initially distributing and visualizing the data. This paper reduces the computational complexity of MCNN, a rigorous data pre-processing technique, which is implemented to enhance data quality, tune hyper parameters, and optimize the feature selection process. Finally, the MCNN model is trained on the training data, tested on the test data, and its efficiency is verified in terms of prediction accuracy separately for spatial, temporal, and spatiotemporal features. The output of MCNN is compared with that of existing learning algorithms to evaluate its performance. Overriding, overruling, sudden stopping, illegal turning, no-entry entering, accidents, and disturbing other vehicles are the anomalous traffic behaviors considered in this paper. From the experimental comparison, it is found that the MCNN outperforms the different models.

Keywords: Anomaly Detection, Road Traffic Abnormalities, Multi-CNN, CCTV Camera, Road Surveillance Monitoring.

INTRODUCTION

Urbanization and rapid population growth have led to a significant increase in vehicular traffic worldwide, posing serious challenges to traffic management authorities, urban planners, and public safety systems. With road networks becoming increasingly congested and complex, ensuring real-time monitoring and swift response to anomalies, such as accidents, illegal parking, traffic violations, and road obstructions, has become a critical priority [1]. Traditional traffic surveillance systems, although equipped with Closed-Circuit Television (CCTV) cameras, often rely on human operators to manually observe, interpret, and respond to incidents. This manual approach is inherently limited by human attention span, prone to errors, and ineffective in scenarios requiring round-the-clock monitoring or rapid decision-making [2]. Therefore, there is a pressing need for automated, intelligent, and real-time traffic anomaly detection systems that can operate efficiently and accurately at scale.

Recent advancements in computer vision and deep learning (DL) have shown great potential in automating the analysis of video surveillance data. In particular, convolutional neural networks (CNNs), a subset of DL architectures, have achieved state-of-the-art results in object detection, tracking, and scene understanding tasks [3]. CNNs are capable of learning spatial hierarchies of features from visual data, enabling them to distinguish between normal and anomalous activities based on contextual cues such as motion patterns, object trajectories, and interactions. However, most existing DL-based traffic monitoring systems are deployed in cloud environments, which introduce several critical limitations. These include high latency due to network delays, data privacy concerns during video transmission, and scalability challenges in areas with limited connectivity or bandwidth [4].

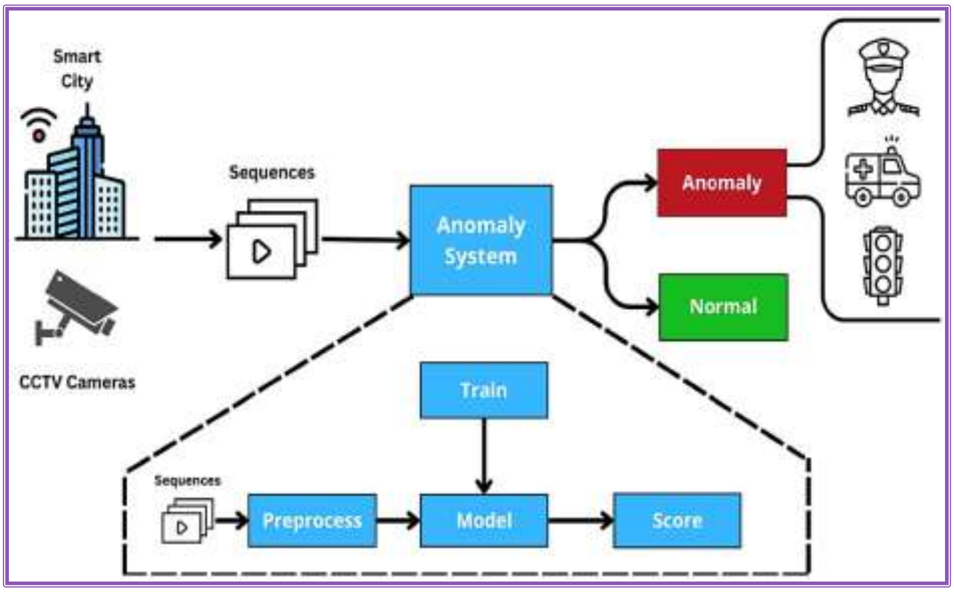

Figure 1. Anomaly Detection Using CCTV Cameras

To address these challenges, the focus of this research shifts toward edge computing, a paradigm that brings computational power closer to the data source—i.e., the surveillance cameras. By deploying DL models directly on edge devices (such as NVIDIA Jetson, Raspberry Pi with TPU, or AI-optimized microcontrollers), traffic video feeds can be analyzed in real-time, without the need to stream data to centralized servers. This not only reduces latency and bandwidth usage but also enhances privacy, responsiveness, and fault tolerance of the system. The adoption of edge-deployed AI marks a significant step forward in enabling smart city infrastructure that is autonomous, efficient, and context-aware [5]. This study proposes a novel multi-CNN-based architecture optimized explicitly for edge deployment to detect traffic anomalies from live CCTV feeds. Unlike monolithic models, the multi-CNN framework enables modular processing of various aspects of the scene—such as vehicle behavior, pedestrian activity, motion irregularities, and spatial violations—by combining the strengths of multiple lightweight CNN models [6, 7]. Each model is trained on a specific task (e.g., vehicle detection, behavior classification, collision detection), and their outputs are aggregated using a decision fusion mechanism to infer anomalies with high precision. The system is designed to run on edge hardware with constrained computational resources, necessitating optimization techniques such as model quantization, pruning, and hardware-aware neural architecture search (NAS) [8].

The dataset for this research comprises a combination of publicly available traffic video datasets, such as the AI City Challenge and UCSD Anomaly Dataset, as well as custom-curated CCTV footage annotated for real-world events [9]. The model's performance is evaluated using standard metrics, including accuracy, precision, recall, F1-score, inference time, and frame-per-second (FPS) throughput under edge deployment conditions. Special emphasis is placed on the real-time responsiveness, scalability across diverse urban environments, and robustness to lighting, weather, and occlusion variations commonly encountered in outdoor surveillance. Furthermore, the research incorporates explainable AI (XAI) tools to visualize how individual CNNs contribute to the anomaly detection pipeline, thereby fostering transparency and interpretability, critical factors for adoption by urban authorities and traffic management agencies. The proposed edge-deployed solution not only enhances situational awareness but also enables proactive measures such as automated alerts, integration with traffic signal systems, and coordination with emergency response units [10].

In summary, this study presents a comprehensive framework for real-time traffic anomaly detection using a multi-CNN deep learning model deployed on edge devices, aiming to revolutionize the way urban traffic systems are monitored and managed. By leveraging the synergy between deep learning and edge computing, the proposed system ensures low latency and high accuracy in surveillance, while addressing the practical limitations of existing cloud-based solutions. This research contributes to the advancement of intelligent transportation systems (ITS), laying the groundwork for safer, smarter, and more responsive urban mobility ecosystems. This paper proposes an advanced real-time traffic anomaly detection system utilizing a multi-CNN architecture installed in edge computing devices, combined with a CCTV surveillance camera system. The major contributions are:

- The various traffic anomalies are detected using an innovative approach that combines an ensemble of CNNs, which identify anomalies such as accidents, congestion, wrong-way driving, and illegal parking with greater accuracy and speed from real-time CCTV footage.
- This model minimizes the requirement of cloud dependencies and network bandwidth,

which ensures low-latency bandwidth directly, and the proposed model is more effective for edge devices

- This model can process live video using a lightweight, robust detection pipeline and quickly alert and respond in traffic monitoring systems.
- The proposed model, compared with traditional models such as single-CNN or cloud-based models, achieves superior performance metrics, including accuracy, latency, and F1-score, as demonstrated by experimental results.
- The proposed system can handle more anomaly types without preserving the whole network, and it can be easily combined with existing smart city infrastructure

Literature Survey

A detailed survey is conducted to understand the issues and challenges faced by earlier methods, which leads to the creation of a pathway for designing an innovative proposed model for analyzing spatiotemporal data to predict anomalies in a road traffic environment. For example, Cheng et al. (2015) developed an early deep learning approach for anomaly detection in crowded scenes using sparse reconstruction techniques. The core idea was that normal patterns in surveillance videos could be efficiently reconstructed from a learned dictionary, whereas anomalies would yield higher reconstruction errors. Although not explicitly designed for edge computing environments, this study laid the foundation for understanding motion irregularities in video surveillance by leveraging both appearance and motion features. The approach, however, was computationally intensive and suffered from limitations in handling complex real-time scenarios, making it impractical for deployment in latency-sensitive and resource-constrained edge environments. Sabokrou et al. (2017) introduced a fast and deep video anomaly detection framework using 3D Convolutional Neural Networks (CNNs), which enabled the model to learn spatial-temporal features directly from video inputs. Their framework was able to operate in near real-time and produced promising results in detecting unexpected events in crowded scenes. The use of fully convolutional architectures allowed for better generalization across different scenes and backgrounds. Despite its real-time inference capability, the method still relies on centralized servers or high-performance GPUs, which limits its feasibility for real-time inference at the edge, where computational resources are scarce. Liu et al. (2018) proposed the *Future Frame Prediction* approach, which detects anomalies by forecasting what the next frame in a video should look like using ConvLSTM architectures, and then comparing the predicted frame to the actual captured frame. Significant discrepancies between the two indicate possible anomalies. This method eliminated the need for labeled anomaly data during training, thus operating in an unsupervised manner. While the prediction-based technique achieved good accuracy in various traffic surveillance scenarios, its high memory consumption, increased inference latency, and the need for recurrent computations made it unsuitable for edge-based real-time implementations. Sultani et al. (2018) made a significant contribution to the field by introducing UCF-Crime, one of the most extensive publicly available datasets for anomaly detection in surveillance videos. Their deep MIL (Multiple Instance Learning)-based framework worked by learning to detect abnormal segments in untrimmed videos. While this allowed the model to understand complex behaviors and achieve state-of-the-art accuracy, its high training and inference complexity, combined with the need for robust cloud infrastructure, limited its application in real-time edge-based CCTV systems, especially in bandwidth-constrained environments.

Ravanbakhsh et al. (2019) presented a hybrid approach that merged spatial features extracted through 2D CNNs with temporal dependencies captured using LSTMs. Their model was particularly suited to capturing abnormal vehicle behaviors, such as sudden stops or lane violations in highway videos. The architecture was efficient and showed potential for real-time inference. However, its memory usage remained a concern for edge devices, as the LSTM components often led to bottlenecks in terms of speed and scalability on hardware with limited computational resources. Tang et al. (2020) proposed a two-stream CNN architecture that processed both appearance and motion information in parallel to identify anomalies in traffic surveillance. Their model employed lightweight convolutional layers and was explicitly optimized for embedded systems such as Raspberry Pi and NVIDIA Jetson Nano. This work addressed both the computational and power constraints of edge devices, making it more feasible for deployment in real-world smart surveillance scenarios. The study also explored real-time optimizations, such as pruning and quantization, to enhance edge inference capabilities without compromising detection accuracy.

Ishak et al. (2021) further advanced edge-based anomaly detection by deploying YOLOv3 alongside a lightweight CNN on the NVIDIA Jetson TX2. Their system was capable of detecting traffic incidents, such as accidents, illegal turns, and stopped vehicles, in real-time. Emphasis was placed on early detection and power efficiency, both of which are critical for long-duration operation in outdoor edge environments. This research highlighted the practical integration of object detection with anomaly recognition on constrained edge hardware, emphasizing modularity for adapting to various urban monitoring needs. Nguyen et al. (2022) addressed privacy and communication bottlenecks in traffic surveillance systems by proposing a federated learning-based framework. Their model distributed CNNs to edge devices across CCTV nodes, enabling local learning and reducing reliance on cloud aggregation. The federated setup not only protected data privacy but also reduced latency and communication overhead. Their work demonstrated that combining federated architectures with lightweight CNNs could achieve both real-time performance and robust anomaly detection, making it a viable model for decentralized traffic surveillance systems. Xu et al. (2022) focused on constructing a strong and lightweight multi-CNN architecture by integrating MobileNetV2 and ResNet18. This ensemble allowed the system to balance inference speed and detection accuracy while consuming low computational resources. The model was successfully deployed on various edge devices like NVIDIA Jetson Nano and Google Coral. Additionally, the incorporation of adaptive thresholding helped minimize false alarms in urban environments characterized by high variability and frequent occlusions. This research validated the potential of hybrid CNN architectures for practical, real-time deployment at surveillance edges. Patel et al. (2023) adopted a holistic approach by developing a comprehensive, end-to-end real-time traffic anomaly detection pipeline utilizing edge AI solutions. Their multi-CNN ensemble system was optimized for Intel Movidius and Jetson Nano modules. The models were trained to recognize a wide variety of anomalies, including illegal U-turns, sudden halts, collisions, and jaywalking. The system employed model parallelism and pipeline partitioning to balance the workload across edge processors, demonstrating that carefully designed CNN ensembles can rival cloud-based systems in both accuracy and speed when deployed with hardware-level optimizations.

Limitations and Motivations

Although the proposed multi-CNN edge-based system offers numerous benefits in terms of speed and accuracy, it also has several shortcomings. To start, hardware limitations of edge devices can impose restrictions on model complexity and the capability to process high-resolution video or multiple video streams in parallel. This may affect the precision of detection in dense areas or visually demanding setups. Second, the generalization of models is also an issue, as rare or unobserved variables in traffic cannot be adequately identified unless sufficient representative data is presented during training. Third, poor environments, such as inadequate lighting, heavy rain, or obstructions, can deteriorate the video quality, resulting in misclassification or failure to detect. Additionally, real-time performance is susceptible to ideal frame rates and camera locations, which cannot be standardized in various urban environments. Finally, there should also be no centralized data aggregation that would constrain the system-scale learning or improvement without regular updates or at-scale retraining.

The primary objective of this research is to increase interest in enhancing traffic monitoring within cities and ensuring the safety of the urban population by introducing automated, smart mechanisms. With the widespread adoption of CCTV systems in smart cities, the possibility of converting passive, security-related surveillance into proactive incident identification is also high. The manual approach is rather resource-demanding and can cause delays, whereas the cloud-based approach may introduce latency and create privacy issues. Hence, the solution to this problem may come about through the usage of DL on edge computing, specifically multi-CNN models. This enables the detection of anomalies in real-time, right where they occur, which reduces the response time and makes the traffic authority instantly aware of the situation. It is also motivated by the idea of developing scalable, low-cost solutions that can operate independently with minimal human interference in finding solutions to a wide range of urban environments. The result of all this is the development of a robust, real-time traffic monitoring and control architecture that considers privacy in the context of smart city applications, thereby enhancing the overall resilience and security of urban environments.

Problem Statement

Rural towns are evolving and shifting towards urban environments, ensuring public safety through a real-time monitoring and management system specifically designed for traffic. Even though the CCTV surveillance system is installed, most of the urban traffic system still relies on manpower to detect anomalies such as congestion, accidents, and rule violations. The surveillance relies on manual oversight, leading to delays in incident detection, improved traffic flow management, and a high risk of secondary accidents. Moreover, several challenges exist in Cloud-based AI solutions, including network dependency, high latency, and data privacy issues. So, the proposed real-time traffic anomaly detection system demands low latency and is automated by being deployed directly at the edge. The challenges in developing a model for a real-time traffic management system lie in achieving accuracy, designing a lightweight system, and analyzing continuous video streaming from CCTV. The gaps in real-time traffic anomaly detection are addressed, and this proposed multi-CNN edge-deployment solution acts as a bridge to resolve the challenges.

Proposed Model

The proposed system involves real-time monitoring of traffic anomalies, utilizing an edge computing device integrated with CCTV surveillance through an advanced framework, specifically a multi-convolutional neural network (multi-CNN). This framework is designed for continuous streaming of video through edge computing, which reduces latency and minimizes data transmission in centralized servers. The impact of this framework in detecting traffic anomalies includes identifying numerous accidents, illegal vehicle stops, traffic congestion, and wrong-way driving. The utilization of multi-CNN ensures a high accuracy and robust performance under critical conditions such as rainy, lightning, and environmental conditions. These advantages are being executed through edge computing, which enables anomaly detection in traffic and alerts traffic authorities to ensure the overall road safety. The enhancement of this framework involves a decentralized approach to data privacy, system scalability, and reduced bandwidth usage, providing an efficient system for monitoring applications in city traffic.

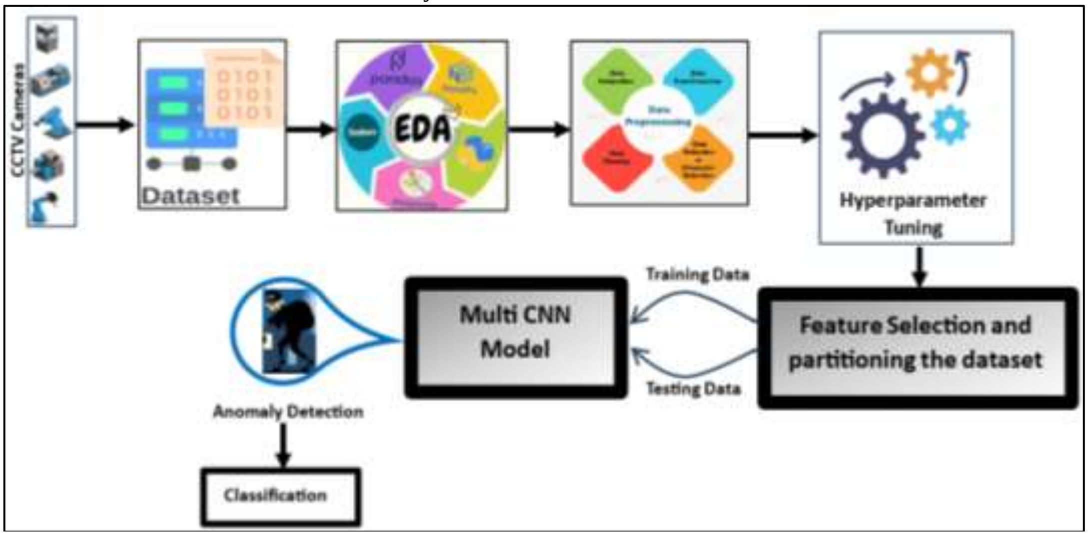

Figure-2. Proposed Framework

Edge Deployment Optimization

The continuous framework used in multi-CNN architecture integrates edge computation, utilizing quantization and pruning techniques to optimize for edge hardware, such as NVIDIA Jetson and Raspberry Pi, thereby ensuring the system's accuracy. In real-time mortaring, the frameworks utilize parallel thread-based frame streaming to execute inference within the system.

Video preprocessing and frame extraction

Broke down the video into a sequence of frames in a continuous stream through edge computing devices, where f_i is resized and normalized for model compatibility. They determined through an equation:

$$F = \{F_1, F_2, \dots . F_n\}, \quad where\ f_i \in \mathbb{R}^{H \times W \times C}$$

Multi-CNN Architecture

The diagram illustrates a Multi-Channel Convolutional Neural Network (Multi-CNN) model (Figure 3) designed for effective anomaly detection and classification, which processes both spatial domain data and time domain data simultaneously. This dual-input structure enables the model to capture and learn from two distinct yet complementary perspectives of the data, thereby significantly enhancing its ability to detect irregularities. At the input stage, the model receives data from two separate domains. The spatial domain data contains information related to positional or structural attributes, while the time domain data focuses on how these attributes change over time. Each of these data streams is fed independently into two parallel convolutional neural network pipelines. Each pipeline begins with a 2D convolutional layer using a 3×3 kernel and 32 filters. This layer helps extract local features by scanning over the input data and identifying spatial or temporal patterns. After convolution, a max pooling layer with a 2×2 window is applied to each stream. Max pooling reduces the dimensionality of the feature maps while retaining the most significant information, thus improving computational efficiency and promoting generalization. The outputs of both CNN pipelines are then merged to form a joint feature representation. This fusion step is critical because it combines spatial and temporal insights, enabling the model to consider both structural and dynamic characteristics of the data simultaneously. This combined representation is more informative and discriminative for downstream tasks. The fused features are passed into a deep CNN-based feature extractor. This part of the architecture consists of multiple layers, each comprising convolution, pooling, and ReLU activation. These layers progressively extract higher-level and more abstract features that are crucial for distinguishing between normal and anomalous patterns in the data.

Following feature extraction, the high-dimensional representation is flattened and sent through several fully connected (FC) layers with ReLU activation functions. These layers serve to transform the learned features into a compact representation suitable for classification. The final FC layer produces an output indicating whether the input is normal or contains an anomaly. At the end of the pipeline, the system outputs results where anomalies are detected and classified. In the visualization, anomalies are marked with red stars, while blue stars represent normal instances. This final step enables the model not only to flag irregular behavior but also to distinguish between different types of anomalies as needed. This multi-CNN model leverages both spatial and temporal domains to create a rich, unified representation of the input data. By using parallel CNN streams, feature fusion, and deep classification layers, the model achieves robust and accurate anomaly detection, making it highly suitable for applications such as industrial fault detection, system monitoring, and intelligent surveillance.

This architecture is used for detecting a specific anomaly, such as a vehicle collision, for which it has been trained. The output of each CNN model is integrated with a group of models, and is then M = $\{CNN_1, CNN_2, \dots, CNN_k\}$:

$$P_i = Softmax\ (W_i * f + b_i)$$

Where f is the input of the frame, W_i is the weight and bias of i^{th} CNN, and P_i is a probability distribution across detection classes. The output has been combined using a fusion layer and expressed as:

$$P_{finial} = \frac{1}{k}\sum_{i=1}^{k} P_i$$

If an anomaly is flagged in the model, $\max(P_{finial}) > \theta$, where θ is a predefined confidence threshold.

Multi-CNN-Based Anomaly Detection

The proposed anomaly detection framework is based on a multi-CNN for detecting various unusual traffic events, such as illegal parking, accidents, wrong-way driving, and congestion. These unusual traffic events are analyzed through the real-time CCTV footage. This proposed framework groups the CNNs, where each CNN model is individually trained to detect a specific class of anomaly. This multi-CNN model enhances detection accuracy while minimizing false positives, enabling the efficient detection of particular events across multiple anomaly types. Moreover, this enables low-latency detection, which is particularly important for real-time anomaly detection systems. The multi-CNN framework offers several advantages, including specialization, parallelism, and real-time performance.

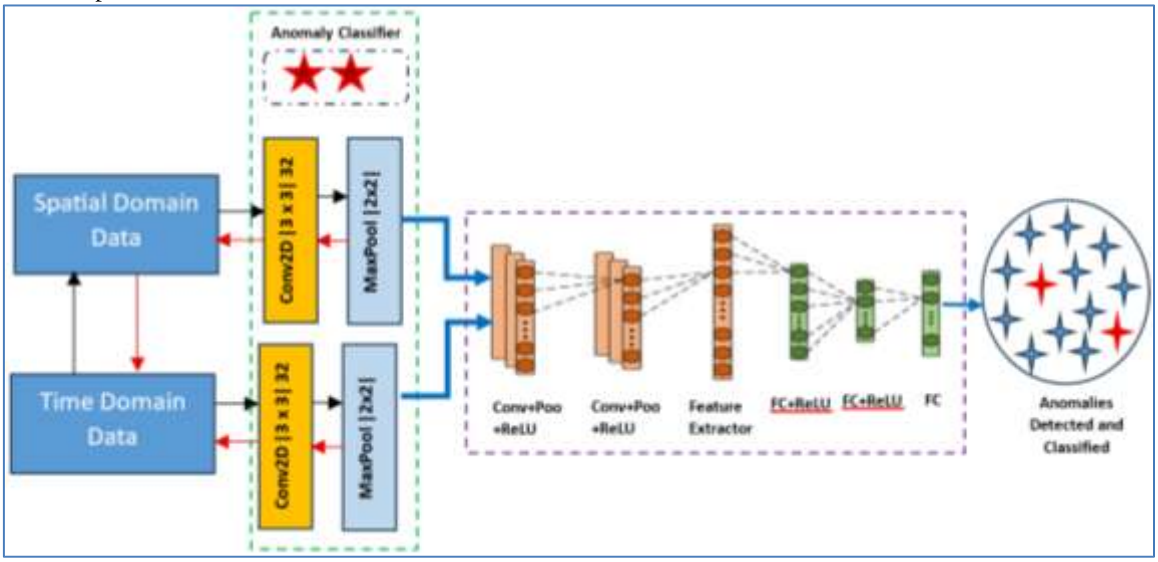

Figure-3. Multi-CNN Model

Frame Preprocessing

The CCTV video streams are segmented to obtain the image frames:

$$F = \{f_1, f_2, f_3, \dots, f_n\}, \quad f_i \in \mathbb{R}^{H\times W\times C}$$

Where the height, width and channels of the frames are denoted as H, W and C. Each frame f_i is normalized and resized for the model input.

Multi-CNN Architecture

The k number of CNN are collected and denoted as $M = \{CNN_1, CNN_2, \dots, CNN_k\}$. Each CNN model is particularly trained for a specific anomaly type. Each CNN function, the CNN_i is applied on input frame f:

$$P_i = softmax\ (W_i * f + b_i)$$

Where the class probability of each CNN_i output is denoted as P_i, convolutional operator is denoted as *, the weight and bias parameter are denoted as W_i and b_i.

Each CNN extracts the features map performed a convolutional and pooling process:

$$Y^{(l)} = \sigma\left(W^{(l)} * Y^{(l-1)} + b^{(l)}\right), \qquad l = 1, 2, 3, \dots, L$$

Where the number of layers and ReLU activation function is denoted as L and σ, and $Y^{(0)}$ is equal to the frame f.

3. Fusion and Decision Logic

Either average fusion or majority voting methods are used to combine all CNN output $\{P_1, P_2, P_3, \dots, P_k\}$.

The average fusion method is used:

$$P_{final} = \frac{1}{k}\sum_{i-1}^{k} P_i$$

The decision rule method is used:

$$Anomaly = \begin{cases} 1 & if\ \max(P_{final}) > \theta \\ 0 & otherwise \end{cases}$$

Where the confidence threshold is denoted as θ. Let's assume the value of the threshold is 0.7.

Output

The value of the maximum probability P_{final} is more than the threshold value θ. The proposed multi-CNN-based system identifies and flags the corresponding anomaly type, and also generates a real-time alert. The object detection extensions are used to overlay the bounding box and label on the frame in real-time CCTV footage.

Data pre-processing

Data pre-processing has created an impact on noise, lighting variation, and occlusion in CCTV footage. The raw video frame F from the continuous streaming is resized to a fixed resolution and normalized to the range. Gaussian mixture models are used to isolate background moving objects using an equation:

$$F_{clean} = F - B \text{ , } where\ B = background\ model$$

Temporal smoothing and frames are used for differentiating the moving regions in the video frame. Optical flow is used to compute and extract motion vectors $V(x, y)$ from a frame sequence, capturing dynamic scene changes.

Feature extraction

The feature extraction in the input image $I \in R^{p \times p \times 3}$, which divides into non-overlapping patches $P_i \in R^{p \times p \times 3}$ using a lightweight vision transformer, spatial and temporal. Every patch is linearly embedded into a vector:

$$z_0 = [x_p^1 E, x_p^2; \dots ; x_p^n E] + E_{pos}$$

Where E is the learnable matrix and E_{pos} is positional encoding. The linear embedding's are passed through multi-head self-attention (MHSA). This allows for capturing the spatial region due to the capability to capture the long-range dependencies across patches and anomalies, when there is a change in motion direction through:

$$MHSA(Q, K, V) = Concat(h_1, \dots, h_h) \text{ ,}$$

$$h_i = Attention(QW_i^Q, KW_i^Q, VW_i^V)$$

Feature selection

Feature selection is used to reduce the computational load on the edge device and to focus on patterns and attention-based scores in ranked features. Let $f_i \in R^d$ to be extracted vector from patch I and a_i is attention weight. This selection is used for ensuring real-time processing, where the feature with the top-k attention weights ($\propto_i > \tau$) are been retained in anomaly detection through an equation:

$$\propto_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)} \quad , \quad e_i = W^T f_i$$

Anomaly detection using a lightweight ViT framework

This anomaly detection is the final transformer layer classification, where binary cross-entropy loss is used to train this model. This model is capable of optimization of edge deployment by reducing the layer count, parameter size, and applying quantization techniques. It detects accidents and wrong-way driving with high efficiency and minimal latency through a mathematical model:

$$L_{BCE} = -[y \log(\hat{y}) + (1 - y)\log(1 - \hat{y})]$$

where $y \in \{0,1\}$ indicates normal or anomalous, and $\hat{y}$ is the model's probability.

The proposed work, as explained above, has been implemented and experimented with to verify its performance. It is also evaluated by comparing its results with those of other similar models.

RESULT AND OBSERVATION

The experimental setup comprises public and custom CCTV datasets, such as AI City and UA-DETRAC, for training and deploying the proposed multi-CNN model for real-time traffic anomaly detection. Each CNN model is trained to detect specific anomalies, such as illegal parking, congestion, accidents, and wrong-way driving. The transfer learning with the hyper parameters like batch size 32, Adam, and LR is 0.0001, used to fine-tune the models based on VGG16 and MobileNetV2. The deployment process is performed using the NVIDIA Jetson Xavier NX, and optimization is achieved with TensorRT, resulting in approximately 30 FPS. In a real-time smart city environment, the performance of the proposed model is validated using evaluation metrics such as accuracy, F1-score, Latency, and resource usage.

Table 1: Evaluation Results Of Various Attacks

Action	Precision	Recall	F1-Score
Abuse	1.00	0.95	0.97
Arrest	0.93	0.91	0.92
Arson	0.98	0.88	0.93
Assault	1.00	0.86	0.92
Burglary	0.99	0.93	0.96
Explosion	1.00	0.93	0.96
Fighting	0.97	1.00	0.99
Normal	0.97	1.00	0.98
Road Accidents	0.97	0.99	0.98
Robbery	0.92	0.94	0.93
Shooting	0.97	0.89	0.93
Shop Lifting	0.98	1.00	0.99
Stealing	1.00	0.93	0.96

Vandalism	1.00	0.94	0.97

Table 1 illustrates that the proposed multi-CNN model detects multiple actions with greater precision, recall, and F1-score. High performance is addressed across all classes, with crucial anomalies such as assault, abuse, and vandalism, for which the score is 1.00. Actions such as shoplifting, fighting, and normal traffic have a recall value of 1.00, ensuring that there are fewer missed detections. Most of the actions have an F1-score of around 0.93, indicating a solid balance between precision and recall. The achieved performance metrics represents the effectiveness of the proposed model, and the model also detects a wide range of traffic-related violations in real-time, with minimal false positives or negatives.

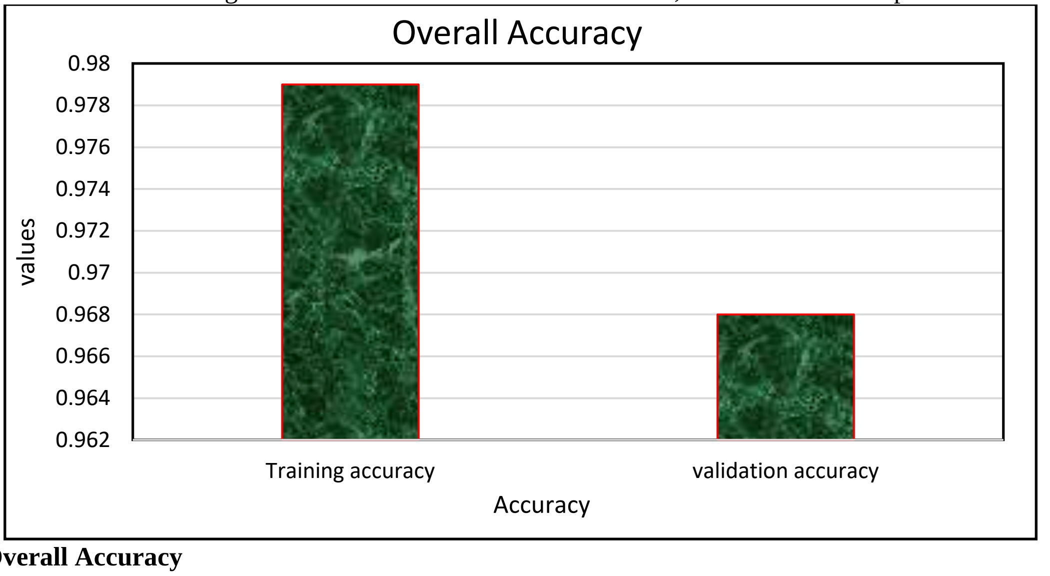

Figure 4: Overall Accuracy

Figure 4 illustrates the overall accuracy progression of the proposed model for anomaly detection at training and validation. The increase in the number of training epochs results in both training and validation accuracy continuously increasing, with the training accuracy attaining around 97.9% and the validation accuracy 96.8%. This close accuracy indicates that the proposed Model effectively learning the patterns with minimal overfitting. The traffic anomalies are accurately detected in real-time with more robustness and reliability, which is assured by the consistent performance in both model training and validation.

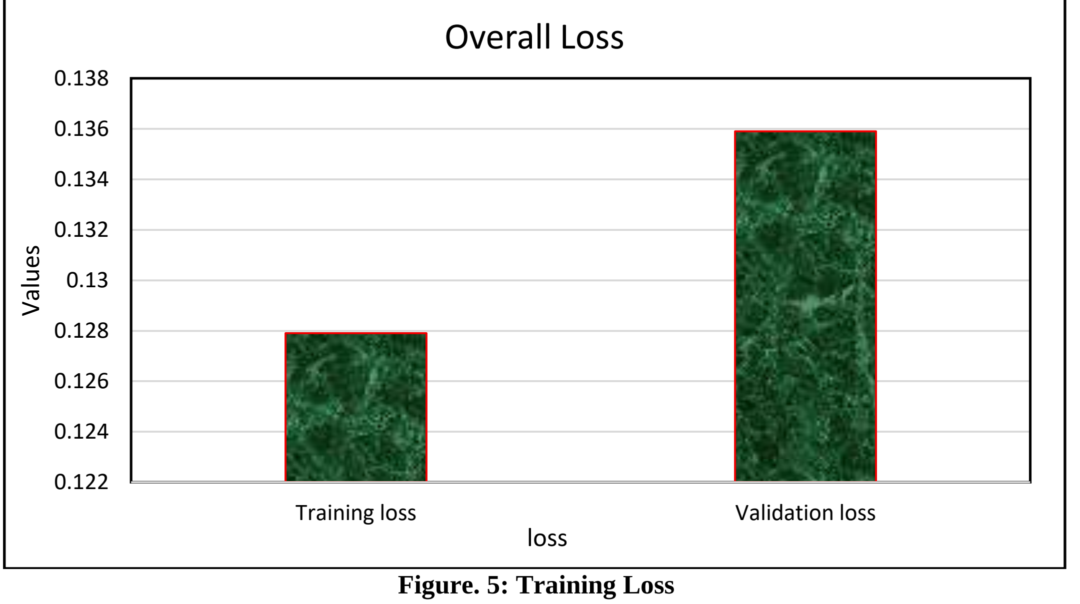

Figure. 5: Training Loss

Figure 5 illustrates that the proposed model has an overall performance loss in real-time anomaly detection. The training loss is around 0.128, which is less than the validation loss of around 0.136. This difference in loss value indicates that the proposed model effectively detects anomalies in new data, with reduced overfitting. Both training and validation losses are too low, indicating that the model is learning effectively and performing reliably. Additionally, the proposed model is suitable for real-time deployment on edge devices for detecting anomalies in traffic using CCTV surveillance footage.

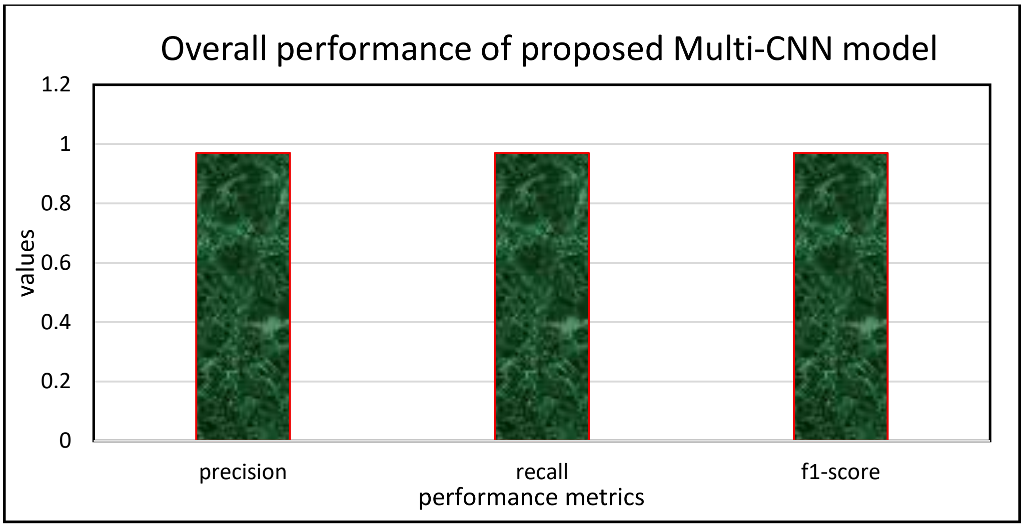

Figure. 6: Proposed Multi-CNN Model Overall Performance

Figure 6 represents the overall performance of the proposed model. It evaluates overall performance through several key metrics, including F1-score, Precision, and Recall. The figure shows that each of these metrics has attained similar results. The results show that the multi-CNN model achieves a precision of 0.97, an F1-score of 0.97, and a recall rate of 0.97. These results indicate that the proposed multi-CNN model has performed well in detecting anomalies with a low error rate. Thus, this model is more suitable for real-time anomaly detection using CCTV footage.

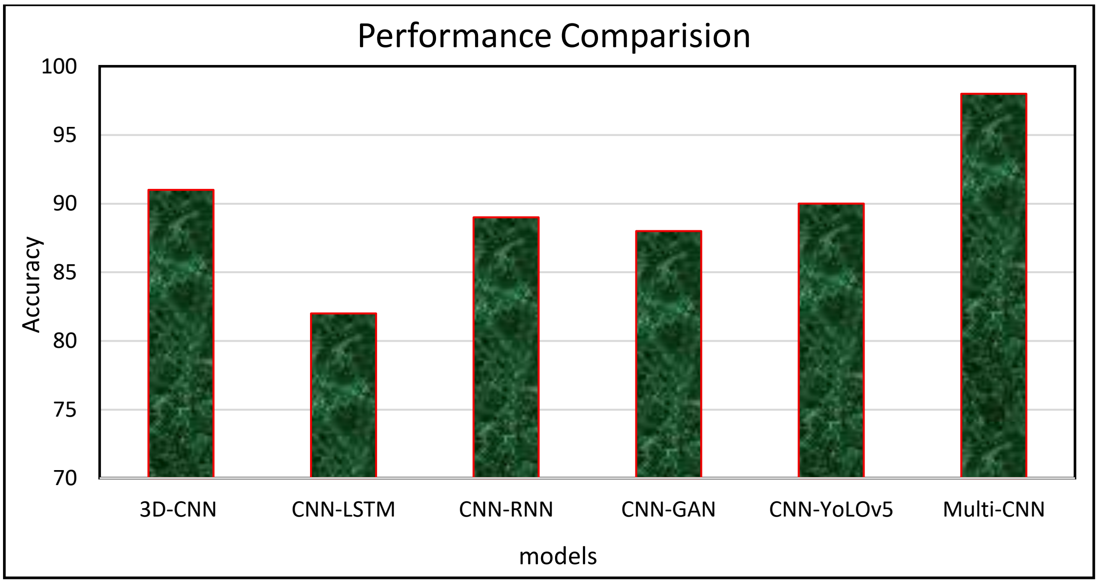

Figure 7. Proposed Multi-CNN Model Comparison

Figure. 7 represents the overall performance comparisons for anomaly detection using CCTV footage. It compares the existing model's performance with that of the proposed multi-CNN model to assess the overall effectiveness of the proposed model. It compares a multi-CNN model with several other models, including CNN-LSTM, 3D-CNN, CNN-RNN, CNN-YOLOv5, and CNN-GAN models. The figures show that the CNN-LSTM model achieved the lowest accuracy rate, detecting anomalies with 82%. The second lowest was achieved by the CNN-GAN model, which attained an accuracy rate of 88%. The 3D-CNN model achieved a detection accuracy of 91%, which is the highest among existing models. At the same time, the proposed multi-CNN model has achieved an accuracy rate of 98% in detecting anomalies using CCTV footage, which is the highest among all models. These results indicate that the proposed multi-CNN model outperforms the existing model in detecting anomalies.

DISCUSSION

This proposed work focuses on a scalable framework for detecting traffic anomalies in real-time, where a multi-CNN model is integrated with edge computing devices connected to a CCTV surveillance network to achieve high-accuracy outputs. This model ensures low latency and efficient bandwidth for enhancing data privacy in managing the local video feed from the system. This architecture combines multiple CNNs to enhance detection accuracy for various traffic anomalies, while also improving operability on edge hardware such as NVIDIA Jetson and Raspberry Pi.

This framework evaluates across multiple traffic anomalies and exhibits better performance over a traditional cloud-based system in terms of detection accuracy, latency, and installation cost. This approach towards more proactive traffic management provides a faster response and contributes to the advancement of smart cities by enhancing the transportation system. This intelligent surveillance monitoring can also be used in the medical field to treat many diseases. They can be applied to monitor cardio vascular diseases particularly. Rare vascular diseases encompass a variety of conditions affecting blood vessels with some notable examples including Buerger's disease, Takayasu arteritis, and Popliteal entrapment syndrome.

.

Future Work

These conditions are often underdiagnosed due to their rarity and the overlap of symptoms with more common diseases, making awareness and research crucial for effective management and treatment. Also this application can extend for more vascular devices to maintain NORD Rare disease database. The future improvement in this framework, with the integration of temporal models such as LSTM and 3D CNN, will provide better insight into motion patterns across video frames, leading to high accuracy in anomaly detection. Another integration of multi-sensor data, such as GPS and LIDAR, supports the visual input and improves detection in visualization in critical environmental conditions. These differ in terms of scalability, where a federated learning framework can be initiated to allow edge devices to learn without transmitting raw video frames, thereby enhancing privacy in continuous streaming for the detection model. Additionally, a self-mechanism for retraining is implemented using supervised learning techniques to adapt to new anomalies in the model. The real-time implementation across traffic zones, with adaptive threshold tuning and performance monitoring in the dashboard, would enhance traffic anomaly detection through more intelligent, autonomous, and resilient systems in a smart city on a global scale.

REFERENCES

1. Pérez, B., Resino, M., Seco, T., García, F., & Al-Kaff, A. (2025). Innovative Approaches to Traffic Anomaly Detection and Classification Using AI. Applied Sciences, 15(10), 5520.
2. Mandal, V., Mussah, A. R., Jin, P., & Adu-Gyamfi, Y. (2020). Artificial intelligence-enabled traffic monitoring system. Sustainability, 12(21), 9177.
3. Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent Advances in Deep Learning for Object Detection. Neurocomputing, 396, 39-64.
4. Aminiyeganeh, K., Coutinho, R. W., & Boukerche, A. (2024). IoT video analytics for surveillance-based systems in smart cities. Computer Communications, 224, 95-105.
5. Patrikar, D. R., & Parate, M. R. (2022). Anomaly detection using edge computing in a video surveillance system. International Journal of Multimedia Information Retrieval, 11(2), 85-110.
6. Chen, J., Li, K., Deng, Q., Li, K., & Yu, P. S. (2019). Distributed deep learning model for intelligent video surveillance systems with edge computing. IEEE Transactions on Industrial Informatics.
7. Nazare, T. S., de Mello, R. F., & Ponti, M. A. (2018). Are pre-trained CNNs good feature extractors for anomaly detection in surveillance videos?. arXiv preprint arXiv:1811.08495.
8. Singh, V., Singh, S., & Gupta, P. (2020). Real-time anomaly recognition through CCTV using neural networks. Procedia Computer Science, 173, 254-263.
9. Chang, M. C., Chiang, C. K., Tsai, C. M., Chang, Y. K., Chiang, H. L., Wang, Y. A., ... & Tseng, H. Y. (2020). Ai city challenge 2020-computer vision for smart transportation applications. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 620-621).
10. Doshi, K., & Yilmaz, Y. (2021). An efficient approach for anomaly detection in traffic videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4236-4244).
11. Cheng, D., Liu, Y., Wu, C., & Yang, Y. (2015). Sparse reconstruction for anomaly detection in crowded scenes. Pattern Recognition, 48(4), 1620–1631.
12. Sabokrou, M., Fathy, M., Hoseini, M., & Klette, R. (2017). Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Computer Vision and Image Understanding, 172, 88–97.
13. Liu, W., Luo, W., Lian, D., & Gao, S. (2018). Future frame prediction for anomaly detection–A new baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 6536–6545).
14. Sultani, W., Chen, C., & Shah, M. (2018). Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6479–6488).
15. Ravanbakhsh, M., Nabi, M., Mousavi, H. S., & Sebe, N. (2019). Spatio-temporal anomaly detection in traffic videos. IEEE Transactions on Intelligent Transportation Systems, 20(3), 1086–1097.
16. Tang, Z., Yu, Z., Gao, J., & Wang, Z. (2020). Lightweight two-stream CNN model for traffic anomaly detection in surveillance videos. Journal of Visual Communication and Image Representation, 73, 102920.
17. Ishak, S., Halim, N. D. A., & Ismail, M. H. (2021). Edge-based video surveillance system for traffic anomaly detection using YOLOv3 and CNN. In

2021 IEEE Symposium on Industrial Electronics and Applications (ISIEA) (pp. 175–180).

18. Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Niyato, D. (2022). Federated learning for smart traffic anomaly detection with edge computing. IEEE Transactions on Intelligent Transportation Systems, 23(1), 705–718.
19. Xu, J., Li, Y., & Zhang, H. (2022). Real-time anomaly detection in urban traffic using multi-CNN models on edge devices. Computer Vision and Image Understanding, 221, 103467.
20. Patel, S., Joshi, K., & Shah, R. (2023). Real-time Traffic Anomaly Detection Using Edge AI: A Multi-CNN Ensemble Approach. IEEE Access, 11, 45029–45042.