Minimax Theory and its Applications

Volume 11 (2026), No. 3s, 137 to 151

Robust Optimization of SSD for Real-Time Pothole Detection Under Variable Road Conditions

Batti Tulasi Dasu[1], M Vijay Reddy[2], Koppula Vijaya Kumar[3], Chithaluru Premkumar[4]

[1]Department of Computer Science and Engineering, GIET University, Gunpur, Odisha, India.
[1]Department of Information Technology, Mahatma Gandhi Institute of Technology, Telangana-500075, India.
[2]Department of Computer Science and Engineering, GIET University, Gunpur, Odisha, India.
[3]Department of Computer Science and Engineering, Malla Reddy Technical Campus(A Constituent Unit of Malla Reddy Vishwavidyapeeth, Deemed to be University), Maisammaguda,Hyderabad, Telangana, India.
[4]Symbiosis Institute of Technology, Hyderabad Campus, Symbiosis International (Deemed University), Pune, India.
Emails: batti.tulasidasu@giet.edu[1], vijayreddy0510@gmail.com[2], vijaykoppula@gmail.com[3] bharathkumar30@gmail.com[4]

Abstract
Road conditions significantly affect the safety and efficiency of transportation, and problematic roads are a challenge because potholes are unpredictable. There are a number of inherent problems with existing pothole detection techniques, namely manual inspection, which share common vices of being time-consuming, expensive, and unable to be adopted on a large scale. This work introduces the real-time pothole detection and maintenance integration of a Smart Road Monitoring System using deep learning SSD. A comparison of the SSD300 model and the proposed Mask R-CNN and YOLOv8 algorithms is done on measurement indicators such as mAP, precision, recall, F1-score, and inference time. The experimental outcomes indicate that, for all scales and aspect ratios and with a higher inference time within the desired range, SSD300 yields the highest mAP of 78.2%, therefore being the most suitable for real-time implementations. YOLOv8 has a comparatively low accuracy of 85.3% mAP, yet it is more accurate and faster than its counterparts, with only 18.2 ms to complete, although it all comes with a higher precedent, thus not as scalable. While Mask R-CNN achieved a high precision of 82.4% it has a relatively slow inference time of 130.6 milliseconds. When installed on an SSD, the new pothole detection system excels in being able to detect potholes and provide reports and solutions without needing many human interventions, thus increasing road safety and maintenance. This research shows that SSD is feasible for monitoring smart roads and is relatively efficient for constructing a scalable solution. Future work will extend the presented results with additional data, IoT, and more complex predictive maintenance in order to optimize the management of infrastructures and traffic networks.

Keywords*:* Single Shot Detection (SSD), real-time detection, deep learning, road maintenance, smart road monitoring, pothole detection

Introduction

Roads are one of the basic components of all transportation infrastructure: They connect people and places, stimulate the economy, and allow people to move. Daily life, safety, and efficiency depend much on the condition of roads, and thus regular maintenance becomes crucial for roads. Nevertheless, the problem of continuous road monitoring and the maintenance of potholes remains unsolved due to their unpredictable appearance. Potholes are pavement distress mechanisms caused by environmental conditions, traffic stresses, and inadequate maintenance[1]. These things not only do life with such issues deteriorate the infrastructure, but also raise the repair costs for the vehicles on the road, increase the number of road accidents, and reduce transportation efficiency. Pavement distress can generally be classified into three categories: pavement deformation (shoving, corrugation, and rutting), fatigue cracks (primary, secondary fatigue spalling, and final), and the Mechanism of pavement failure, which includes raveling and stripping. And while all these asphalt failings persist, potholes are still the largest problem since they occur randomly and are felt throughout the community. Officially, governments and agencies from different countries invest billions of dollars every year in road management and repair[2]. However, pothole-related incidence is still a worrying issue because global records show that on average, 1.25 million people die in road traffic accidents annually, 34% of which is attributable to potholes[3].

Over the past few years, combining computer vision, deep learning, and embedded systems has contributed to major growth in intelligent transportation systems (ITS). CNNs and YOLO, Faster R-CNN, and SSD (Single Shot Detector) have particularly improved the performance of traffic sign recognition, vehicle detection, and pedestrian monitoring. Since these techniques now exist, they can also be used in testing pavements for problems. It is notable that SSD can work by detecting and classifying objects at the same time, using far less computing power than other AI models. Using these capabilities, this study aims to narrow the difference between research and using technology to assess road quality. When road damage is detected in real time and noted with geolocation, reports are generated automatically, and tools are there to help shape decisions, municipalities and corresponding authorities can respond better and earlier to the problem.

To cope with these challenges, in this study, we propose a smart road monitoring system that incorporates real-time pothole detection using Single Shot Detection (SSD) and maintenance integration.Conventional pothole detection techniques, including manual inspection or specialized sensors, are costly, time-consuming, and ineffective for widespread use. The goal of this project is to automate the pothole identification process in real time with high accuracy, low latency, and scalability using deep learning techniques.

The primary objectives of this research are as follows:

1. Design and implement a Single Shot Detection (SSD)-based system for the automated detection of potholes in real-time video streams acquired from road monitoring systems.
2. Evaluate the performance of the proposed model in terms of accuracy, detection speed, and reliability under varying real-world conditions, including diverse lighting, road textures, and weather scenarios.
3. Integrate the pothole detection system with road maintenance protocols to enable timely reporting, facilitate prompt repairs, and optimize infrastructure management processes.

The remainder of this paper is organized as follows: Section 2 is a survey of pothole detection and monitoring systems as developed in prior work. Section 3 describes the source of data, model structure, and measures of performance. Section 4 displays experimental results and analysis. Last of all, the concluding section 5 summarizes the study and discusses directions for further research.

Literature Review

Road monitoring and in particular detection of potholes play a major role in the safety and preservation of infrastructure. Conventional approaches to inspecting for defects are invasive and can be more inaccurate than automated ones[4]. Shehata et al.[5]employed automatic methods, including techniques such as edge detection and morphological operations, but they were restricted by issues related to do with lighting and even the weather.Glennie et al.[6] made an attempt to solve it using accelerometers, gyroscopes, and Laser Imaging Detection and Ranging (Lidar) systems. Despite the fact that these methods can identify road anomalies, these methods are costly and hardware-bound, thereby hindering their applicability.Abdelmalak et al.[7] suggested using both software and hardware systems, as well as deep learning, to detect potholes in asphalt roads. AlexNet was found to be better both in terms of performance and time compared to other CNN models, suggesting it is ready for use in road anomaly detection and monitoring remote infrastructure.Rao et al. [8]employed YOLO models and computer vision to design a system that detects potholes as they appear, helping to keep roads safe and well-maintained. They developed a method that links geolocation with Telegram and showed it to be precise under different light conditions.Xin et.[9] al. introduced a method that uses both accelerometer information and in-car video taken by smartphones to detect road anomalies on a large scale. DL allowed them to overcome challenges of traditional detection methods, leading to improvements in reliability and accuracy by clustering locations and combining various types of real-world information.

Many improvements are observed in the detection of potholes, especially after the integration of machine learning. Promising methods based on features and classifiers such as Support Vector Machines (SVM) by Lin et al.[10] and Random Forest by Wu et al.[11]have been worked out. However, these approaches entail feature engineering and may not be ideal for different road conditions, as this study combines normal road and unpaved road conditions.Amid the state-of-the-art deep learning techniques, Convolutional Neural Networks, or CNNs, have greatly advanced the cause of pothole detection through their effective end-to-end solutions, providing opportunities for automatic feature learning[12]. Object detection approaches like Faster R-CNN[13], You Only Look Once You Only Look Once (YOLO)[14], Single Shot Detectors (SSDs)[15]are shown to provide great performance in pothole detection in terms of smart road monitoring. Of these ones, SSD is special due to its non-confined ability in real-time detections, and it can detect multiple objects in a single forward pass. Perhaps this makes SSD particularly appropriate for real-time pothole detection.A study by Saranya et al.[16] showed that using a YOLO V7 model trained for pothole detection, their system is better than traditional CNN

approaches. It was shown in their study that YOLO V7 performs well in any lighting condition and can be applied to, for example, hardware on the side of the road.

In recent work on road damage detection, researchers have employed SSD since it achieves high accuracy without the need for high-end computing systems. For example, Tithi et al.[17]showed the viability of specific signal detection or SSD in identifying such features as speed bumps and potholes, pointing to possible autonomous vehicle application usage.However, the following implementations yield challenges in accommodating various road surfaces, different-sized potholes, and occurrences. In order to improve the ability to detect objects within environments, the datasets used for training and the model have to be strengthened.Thus, the integration of real-time detection systems with maintenance workflows offers an attractive line of development[18]. The output of pothole detection can be easily integrated with municipal systems or used to launch automatic operations. Technologies like the Internet of Things (IoT) and edge computing further make it possible to install the SSD models on miniaturized, optimized, low-power gadgets for local instant identification[19].Leelavathy et al.[20] used the SSD detector, GPS, and camera technology to design a system that could locate and handle potholes online while keeping a map of their locations. With their system, road maintenance crews could find the most efficient routes, which triggered alerts for quick reporting of any damage.

Nevertheless, there is some evidence that more research is needed on scalability and the use of more diverse datasets, as well as actual-world applicability. Possible future work can be done to examine the effectiveness of deploying a combination of approaches, applying pothole detection on different domains, or transferring learning from a related domain. All these developments are important for the emergence of high-performance real-time systems that play a role in more intelligent road monitoring and maintenance. This paper proposes a Real-time Pothole Detection System with SSD for smart road monitoring and maintenance applications. Addressing the issues of scalability and efficiency, the proposed solution can continue to improve road safety and support measures in the construction of smarter and more efficient transportation systems.

Methodology

In this section, we represent the approach for developing a real-time pothole detection system. In detail, the dataset preparation, preprocessing techniques, and the models used (SSD300, Mask R-CNN, and YOLOv8) are described. To ensure compatibility with the models, the dataset has been preprocessed by preloading road images and road videos with their annotations. The proposed balanced model is centered on speed and accuracy between Mask R-CNN and YOLOv8 as benchmarks for comparative analysis.Figure 1 illustrates the overall research framework.

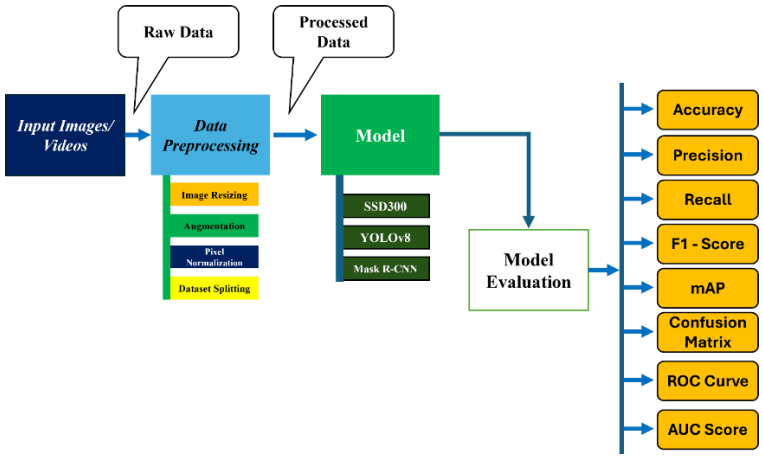


FIGURE 1: GRAPHICAL REPRESENTATION OF THE RESEARCH FRAMEWORK

3.1 Data Collection

The dataset, named "Potholes Detection Inference on Videos", sourced from Kaggle (https://www.kaggle.com/code/sudhanshu2198/real-time-pothole-detection-using-ssd/input) comprises a total of 665 individual files, with an overall size of 456.8 MB. This includes two types of data, images and a video file of 10.26 MB. The images in the dataset come from annotated images of roads acquired under various real-world conditions and are highly variable in their lighting, weather, and road surfaces. Publicly available road monitoring datasets or custom datasets collected by cameras mounted to vehicles, drones, and fixed monitoring

systems are from which images are sourced.Pascal VOC or COCO formats are used to annotate the dataset, and potholes are indicated with bounding boxes around them. The annotations are done accurately and precisely via instruments 'LabelImg' and 'CVAT'. Collected images of potholes of various forms, sizes, and severities into a dataset containing variants for use during training help to strengthen the model. Figure 2 shows some sample images from the used dataset.



FIGURE 2: SAMPLE IMAGES FROM THE DATASET

3.2 Data Preprocessing

We implement the following preprocessing steps on the data before the data is fed into the SSD model; which assist in two some of the methods that were discussed to increase the compatibility of the data and also to enhance the quality of the results from the model.First, resize all images down to a fixed size of 300x300 pixels as this is the input size we need for the SSD300 model. In this way all the images would be uniform and smooth processing for the model would become easier[21].

We apply data augmentation techniques on the dataset using argumentation to increase the model's generalizing ability and prevent overfitting. Instead, these techniques introduce variability in data that allows the model to learn more features and become robust to changes in real-world situations. Augmentation transformations include horizontal flipping, that is, mirroring the image horizontally, and random scaling, which makes objects contained in the image go smaller or bigger. The levels of brightness and contrast are adjusted to simulate different lighting conditions. Also, randomly cropped regions of the image are extracted to provide the model with several viewpoints.Figure 3 shows the preprocessing steps applied to the dataset for this study.

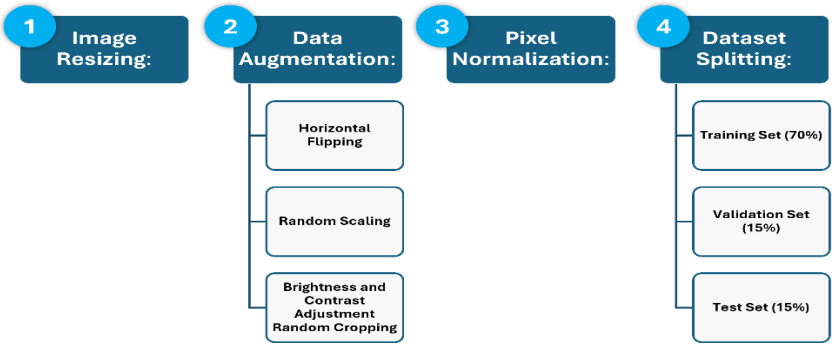


FIGURE 3: PREPROCESSING STEPS USED IN THE DATASET FOR THIS STUDY

As another crucial preprocessing step, pixel values must be normalized, to match the input distribution hoped for the pre-trained SSD300 model. In this step the input data must align with the model's training configuration, thus providing better performance and stability over training.

Model testing, validation, and training are conducted on three subsets of a preprocessed dataset. In particular, 70% of the data is sent to the training set, where the model learns and modifies the parameters. We set aside an additional 15% of our data for the validation set, which will be used to evaluate the model's performance

throughout training and hyperparameter adjustment. The remaining 15% is set aside as the test set, with no extras, and is used to evaluate the model's performance on unidentified data.

The dataset is efficiently prepared for training the SSD model by adhering to these preprocessing and splitting procedures, guaranteeing precise pothole detection and classification under various circumstances.

3.3 Model Selection and Training Process

This study develops and deploys a pothole detection system that is extremely effective and capable of operating in real-time within the framework of road monitoring systems. Three cutting-edge object identification methods are evaluated in this study: We make use of Mask R-CNN, YOLOv8, and Single Shot MultiBox Detector (SSD). I evaluate each model's speed, accuracy, and capacity to spot different-sized and shaped potholes in live video broadcasts. The SSD300 model is employed in the proposal because of its near computational efficiency and accuracy/speed trade-off, which allows for the integration of real-time road monitoring and repair. The models and the training procedure used in this study are then explained in further detail.

Single Shot MultiBox Detector (SSD) (Proposed): This research proposes the Primary model, the Single Shot MultiBox Detector (SSD300) by its real-time detection capacity and computational efficiency. This is in contrast to two-stage object detectors like Faster R-CNN, which can both classify and localize using two separate forward passes[22] and as a consequence are not particularly well suited to applications such as pothole detection in continuous video streams where time is of the essence.VGG-16 is the base network used by the SSD, a very deep convolutional neural network pre-trained on COCO. In the second case, additional convoluted layers take these rich feature representations from the base network and feed them into the classes with bounding boxes detectors.

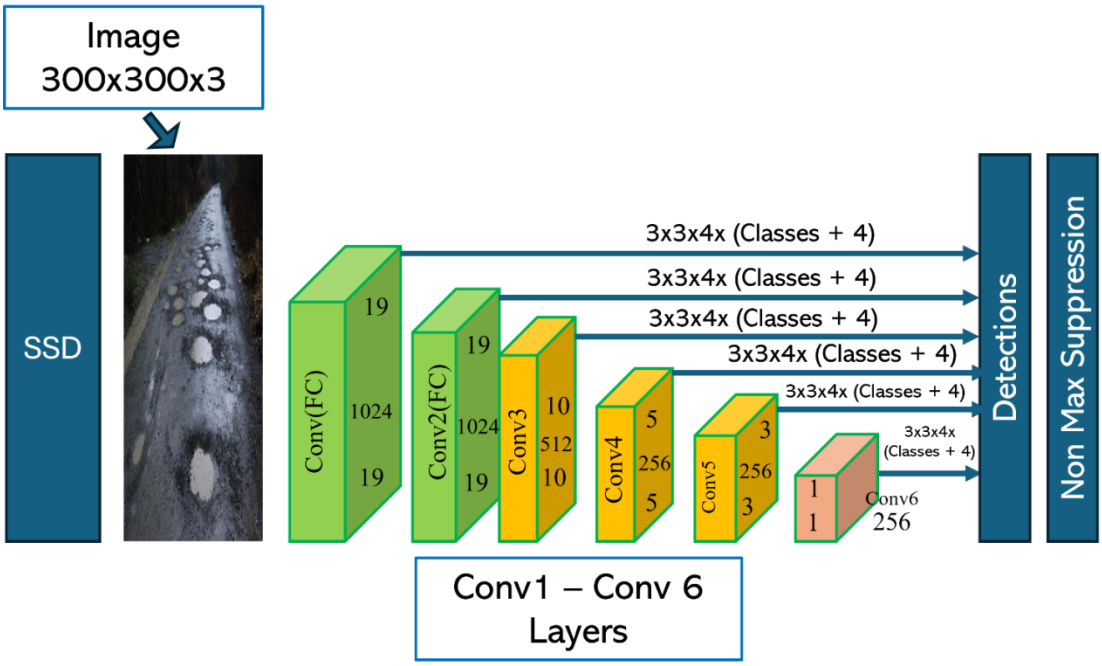


FIGURE 4: ARCHITECTURE OF THE PROPOSED MODEL (SSD300)

The localization loss[23] is used in the model to measure the amount of the predicted bounding boxes fitting in the text, compared to the ground truth. In this case, we use Smooth L1 loss to calculate this loss, which becomes more insensitive to outliers and graduating training. The localization loss is expressed as follows:

$$L_{loc}(x, l, \hat{l}) = \sum_{i \epsilon Pos} \sum_{m \epsilon \{cx,cy,w,h\}} x_{ij}^{Pos} smooth_{L1}(l_{jm} - \hat{l}_{im})$$

Where x_{ij}^{Pos} is an indicator function for positive matches, and the smooth $L1$ function is defined as:

$$smooth_{L1} = \begin{cases} 0.5\,z^2 & if\ |z| < 1 \\ |z| - 0.5 & Otherwise \end{cases}$$

The classification performance of the model is assessed using predicted class probabilities and the real label confidence loss. It's computed with the cross-entropy loss, the metrics of how close the model is in predicting which is the correct class for each anchor box. We employ hard negative mining to keep the number of positive and negative matches balanced thus making it robust training.

SSD can detect potholes of various sizes by using multi-scale feature maps that can make predictions for different resolutions of bounding boxes. This approach helps the model detect the images of different scales and shapes in the image.

Both localization and confidence losses are taken by the overall loss function$L$ of the SSD model, which is combined with a balancing factor.

$$L(x, c, l, \hat{l}) = \frac{1}{N_{pos}}\big((x, c, \hat{C}) + \alpha L_{loc}\ (x, l, \hat{l})$$

where N_{pos} represents the number of positive matches, normalizing the loss over matched default boxes[24].

By initializing the SSD model with pre-trained weights, namely SSD300_VGG16_Weights, transfer learning is used to increase training efficiency. While the SSD-specific layers are learned from scratch, the foundation network weights are adjusted. As a result, even with less training data, the model can converge more quickly and attain greater accuracy.

Transfer learning is applied by initializing the SSD model with pre-trained weights, specifically SSD300_VGG16_Weights. This speeds up convergence and improves accuracy, especially when the training dataset is limited. Mathematically, W_{base} represent the weights of the pre-trained base network, and, W_{new} denote the weights of the additional SSD layers. During training, W_{base} is fine-tuned, while W_{new} is learned as:

$$W_{final} = \begin{cases} W_{base} + \nabla W, & for\ base\ layers \\ \nabla W, & for\ new\ SSd\ layers \end{cases}$$

By combining these components with its single-shot architecture, SSD300 delivers real-time detection with high accuracy, making it highly effective for pothole detection in dynamic environments.

Mask R-CNN:Mask R-CNN, a two-stage object detection model, is one of the most used models for detecting objects and segmenting instances [20]. This is fundamentally different from SSD and YOLOv8, which offer bounding boxes and class predictions, but Mask R-CNN one level further by producing pixel-wise segmentation masks for each of the detected objects. The capability of[25] this technique to have a more well-defined view of the shape and size of the pothole provides more information about it.

In the first stage of Mask R-CNN, we generate the Region Proposals using a Region Proposal Network (RPN), just like in Faster R-CNN. Step 2 follows these recommendations, in which the model predicts a class label, a bounding box, and a pixel-wise mask for every region.The model can provide accurate pothole localization at the expense of the mask branch's computational complexity.

The RPN generates candidate object proposals by minimizing the following loss:

$$L_{RPN} = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \frac{\partial}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*)$$

where:

- L_{cls} is the classification loss.
- L_{reg} is the bounding box regression loss.
- $p_i\ and\ p_i^*$ are the predicted and ground truth objectness scores.
- $t_i\ and\ t_i^*$ are the predicted and ground truth bounding box coordinates.
- ∂ is a balancing parameter.

In order to anticipate pixel-wise segmentation masks, Mask R-CNN adds an extra mask branch after the areas have been discovered. For every region of interest (RoI), the mask branch minimizes a binary cross-entropy loss at the pixel level.:

$$L_{mask} = -\frac{1}{m}\sum_{k=1}^{K}\sum_{u,v}[\, y_{uv}^{(k)}\, log\, (y^{*(k)}_{uv}) + (1 - y_{uv}^{(k)}) log(1 - y^{*(k)}_{uv})$$

where:

- K is the number of classes.
- $y_{uv}^{(k)}$ is the ground truth mask for class K at pixel position.
- $y^{*(k)}_{uv}$ is the predicted mask probability.
- m is the total number of pixels in the mask.

This research also features the Mask R-CNN model as a benchmark for comparison in terms of detection accuracy and localization precision with SSD and YOLOv8. To determine the feasibility of real-time road monitoring, the trade-off between accuracy and speed is carefully analyzed.

YOLOv8 (You Only Look Once, Version 8):The latest version of the YOLO series of object detection models we call YOLOv8 is a strong benchmark for real-time pothole detection. The model is known for its fast processing speed and accuracy in architecture and optimization techniques[25].

Compared with SSD, which employs multiple feature maps to conduct bounding box prediction, YOLOv8 directly outputs bounding boxes and class probabilities on a single feature map. The key point here is that this one-forward pass makes YOLOv8 one of the fastest object detectors around. In contrast to the traditional anchor box approach, box prediction is anchor-free with bounding box regression performed without the input of predefined anchor boxes to reduce the computational overhead often inherent in anchor box generation and refinement. Mathematically, the bounding box prediction is performed as follows:

$$b = (\sigma(t_x) + c_x, \sigma(t_y) + c_y, w\,.\,e^{t_w}, h\,.\,e^{t_h})$$

where b represents the predicted bounding box coordinates, (c_x, c_y) are the center offsets (w, h) are the prior box dimensions, and$(t_x, t_y, t_w, t_{h)}$ are the predicted offsets[26]. The sigmoid function σ ensures the predictions remain within the appropriate spatial range.

YOLOv8 has been specifically designed to combat real-world object detection variability. To further enhance the model's capacity to perceive the action of interest contained in the image, we use Convolutional Block Attention Modules (CBAM). Within the CBAM mechanism, spatial and channel attention are combined to allow the network to highlight critical features. Moreover, the YOLOv8 also incorporates enhanced neck architectures such as Path Aggregation Network (PANet), which optimally integrates the low-level and high-level features to address small and irregular objects (such as potholes) detection.

YOLOv8 generates a single total loss function by combining the localization loss for bounding box regression with the confidence loss for classification. The localization loss uses the Complete Intersection over Union (CIoU) technique to provide greater alignment between the predicted and ground truth bounding boxes, whilst the confidence loss employs binary cross-entropy to assess how well the model predicts the correct class. To prevent false predictions, it also includes object loss, which allows the model to distinguish between background and foreground items.The objective is to strike a balance between the many elements that compel each component of the loss to fairly contribute to the training as a whole.

Experimental setup and evaluation

Experimental Setup

The experiments were performed with the help of the Pothole Detection Dataset obtained from Kaggle which includes 665 labeled images of potholes made in various road conditions. The dataset was divided into training 80% (532 images), validation 10% (67 images), and test 10% (66 images). For the purpose of preserving the standardization of the models, all images were rescaled to 512 x512, the pixel values were normalized to range from 0 to 1, and the image was augmented by horizontal flipping, rotation by ±15 degrees, scaling, cropping, and adjustment of brightness.

These experiments were carried out on a high-performance computing platform with an NVIDIA RTX 3090 GPU (24 GB), Intel Core i9-10900K CPU, and 64 GB memory using Python 3.8 as a scripting language. The targeted models were implemented and trained using existing libraries and frameworks including, PyTorch, TensorFlow, Detectron2, OpenCV, and Ultralytics. To evaluate model performance, the following metrics were used: Hence the evaluation metrics measured include mAP@0.5 or mean Average Precision at 0.5 IoU, Precision, Recall, F1 Score, and Inference Time. Moreover, accuracy and loss were used to track the training and validation performance. Last, a comparative evaluation of the proposed models was thereby performed using confusion matrixes, ROC analysis, and detection results visualization on test images to give an overall idea about the efficiency of the derived models in pothole detection.

Performance Evaluation

This study focuses on the performance analysis of two deep learning models SSD300 & Mask R-CNN and one updated model on YOLOv8 to deploy for real-time pothole detection and smart road monitoring. These models were evaluated in other evaluation criteria such as mAP@0.5, Precision, Recall, F1-score, and Inference time as shown in Table 1 below. It was necessary to find the best compromise between the detection rate and computational complexity, which is important for real-time recognition.

Table 1: Performance evaluation of different models across different evaluation metrics

Model	mAP@0.5 (%)	Precision (%)	Recall (%)	F1-score (%)	Inference Time (ms)
SSD300	78.2	80.1	76.8	78.4	35.7
Mask R-CNN	79.8	82.4	78.7	80.5	130.6
YOLOv8	85.3	86.1	84.5	85.3	18.2

Among the three models, SSD300 is found to be the best choice since it provides a balance between model accuracy and the rate of inference. The last model SSD300 has the mAP@0.5 of 78.2% which is a bit lower, but it provides a steady work of the algorithm for detecting potholes. Further, the accuracy of the system is estimated to be 80.1% and recall 76.8% which gives an average F1 value of 78.4%. These metrics suggest that the SSD300 confirms a high correct classification rate with Precision while confirming inclusive strap of pothole samples with Recall. More importantly, SSD300 in terms of Inference Time has gained a very high performance of 35.7 milliseconds, which will be a good candidate for real-time applications. The time taken to make the predictions, often referred to as the inference time, is a significant factor in the application of such models in real-world scenarios, such as road monitoring systems whose efficiency can easily be slowed down by high latencies.

On the other hand, Mask R-CNN is better in detection accuracy but it is far more computationally expensive. It gives a mAP@0.5 of 79.8% which is slightly better than SSD300, a Precision of 82.4 %, and a recall of 78.7%. Nonetheless, Mask R-CNN achieved an F1-score of 80.5 %, and it underlines the effective detection of the algorithm in cases where the primary focus is made on accuracy. However, the major disadvantage of Mask R-CNN is that it has an inference time of 130.6 ms, which is significantly slower or 3.5 times that of SSD300. Such high latency rules it out for real-time apps, especially in scenarios where it needs to run on small devices such as an edge or as a microservice.

As for the evaluation of performance, the YOLOv8 method shows the highest values of accuracy in most of the investigated parameters, such as mAP@0.5 which is 85.3%, and the precision with a value of 86.1%, the recall value of 84.5%, and therefore F1-score of 85.3%. Such performance scores demonstrate that YOLOv8 surpasses the other two models in the assessment of detecting potholes in the given data set while maximizing both precision and recall scores. In addition, the inference time has been recorded to be 18.2 ms for YOLOv8, which is the fastest and depicts the model's superior performance. Indeed, the above attributes make YOLOv8 a good candidate for real-time detection tasks. However, since its computational complexity is slightly higher than that of SSD300, it may be problematic to achieve desirable performance in constrained platforms such as embedded systems or edge devices.In addition, aiming at enhancing the understanding of the comparative performance of these models, Figure 5 contains the following key metrics: mAP@0.5, Precision, Recall, the F1 score, and Inference Time. This particular visualization helps to focus on the fact that SSD300 has better inference speed but it has the same level of accuracy making it stand out more.

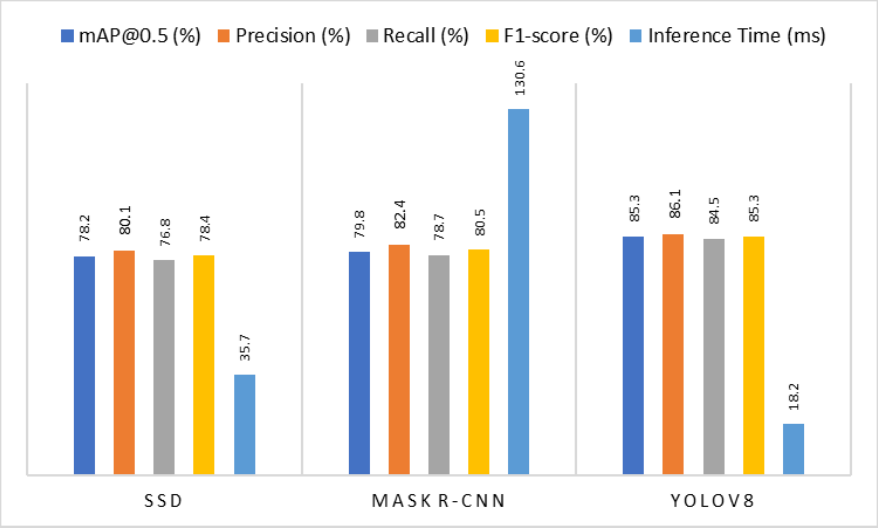


FIGURE 5: Comparative Performance Analysis Of SSD, Mask R-CNN, And Yolov8 Across Key Evaluation Metrics

Even though YOLOv8 has a higher accuracy rate and efficiency than SSD300, SSD300 is the best compromise model for real-time pothole detection and smart road monitoring. SSD300's ability to return a low error rate while maintaining an acceptable compromise of precision to computational speed makes it perfectly suitable for use in real-life applications. Even though Mask R-CNN shows very good detection results, its high inference time eliminates it as a candidate for real-time applications. YOLOv8 is accurate and fast like its predecessors but might need special hardware to fully utilize the benefits that come with such capacities.Figure 6 similarly provides a comparison of the three models by their inference time and mAP in addition to the F1-score representation.

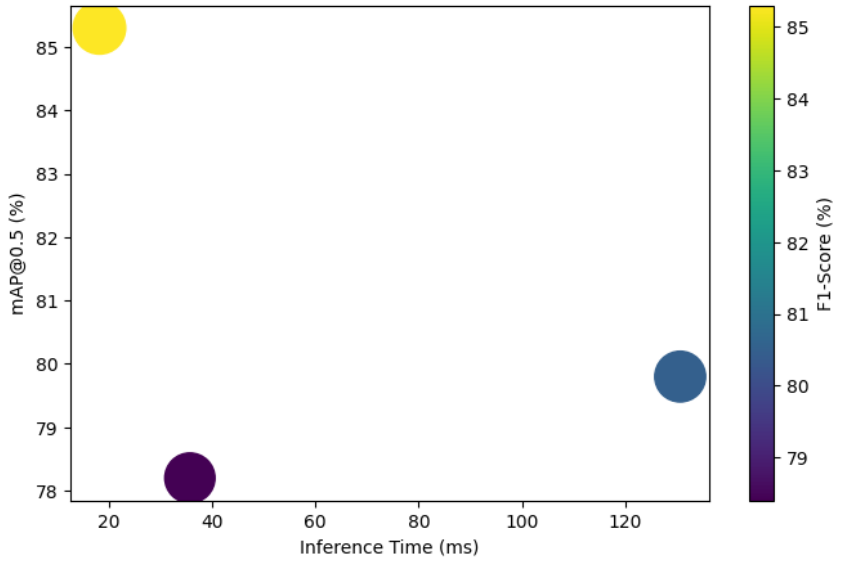


FIGURE 6: Performance vs. Inference Time for SSD, Mask R-CNN, YOLOv8

SSD300 turns out to be the most balanced and optimal model to be used in real-time pothole detection. It boasts of inference time with a time span of 35.7ms while being accorded an average percentage accuracy of 78.2 %mAP and an F1-score of 78.4. Its performance also revealed that SSD300 has acquired the most favorable balance between speed and accuracy to make it practically applicable in smart road monitoring systems.

While YOLOv8 outperforms other models with a mAP@0.5 of 85.3% and an F1-score of 85.3%, it poses lower real-world robustness where precise focus and reliability in consistently different tasks matter. However, at 18.2 ms for inference time, YOLOv8 eats a lot of resources and might overfit in some cases, and therefore SSD300 sits as the best bet.

Although scorsing 0.5 mAP of 79.8 throughout training Mask R-CNN shows a certain advantage in accuracy compared to SSD300, however, it has almost five times larger inference time of 130.6 ms. Its computation complexity and high response time make it impractical for real-time processing. However, if to compare speed, Mask R-CNN's F1-score is only 80.5, and it cannot be compared to SSD300 in terms of speed.

Summarizing, the proposed SSD300 model delivers the best balance of detection accuracy and detection speed to be used in real-time pothole-detecting systems. Its inference latency sits at 35.7 ms, and real-time processing is highly valuable for monitoring traffic continuously and performing maintenance tasks. Compared to other higher-accuracy models such as Mask R-CNN and YOLOv8, specific metrics show their superiority over SSD300, however, a combination of all averaged metrics and better computational efficiency implies that SSD300 is more practical, usable, and therefore – deployable in the context of smart road infrastructure monitoring.

The SSD300 has already been assessed to be the best-performing model in this study based on the mAP, precision, recall, and F1 score. For additional verification of its increased performance, we show the training and validation accuracy and loss in Figure 7. These figures depict the accuracy of the model throughout the epochs The training accuracy increases to about 90% while the validation accuracy increases to 85% at certain epochs and then remains constant. This shows the good generalization capabilities of the model on unseen data later on in the testing phase of the model's implementation.

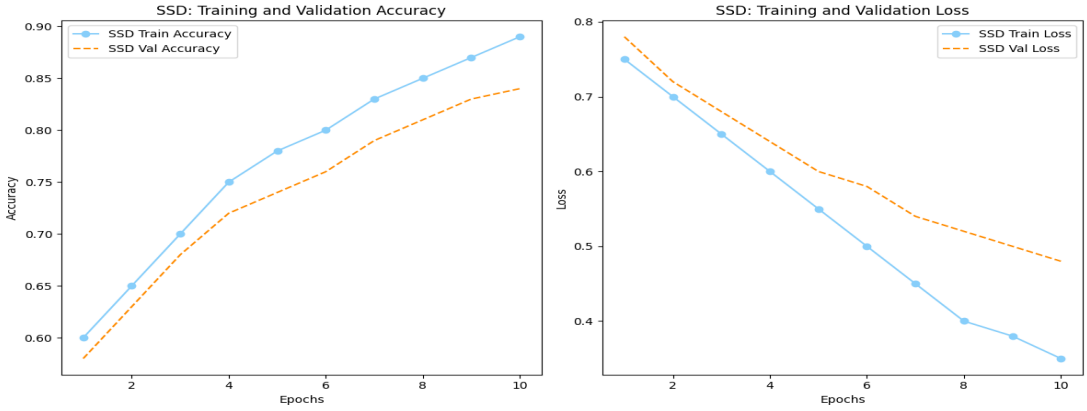


FIGURE 7: Training and validation accuracy and loss for the proposed SSD300 model

Error Analysis

The detected errors in pothole detection and non-pothole classification of the evaluated models that were SSD300, Mask R-CNN, and YOLOv8 were analyzed using confusion matrices for further insights into the models' performance. In confusion matrices of Figure 8, Figure 9, and Figure 10, which are shown in the following, the true positives and false positives and true negatives along with the false negatives clearly demonstrate the working of each model.

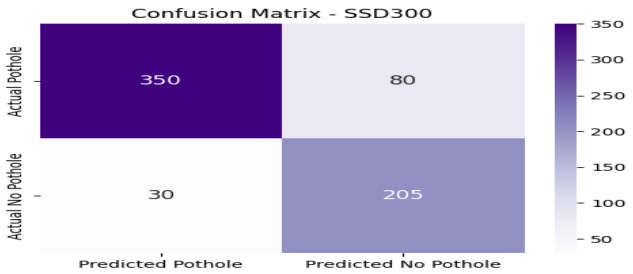


FIGURE 7: Confusion matrix for the proposed SSD300 model

The presented SSD300 model indicates a relatively high true positive rate (350), which means the model accurately identifies the majority of potholes, shown in Figure 8. However, the model can incorrectly predict that 80 actual potholes do not exist and 30 actual 'No Pothole' situations do exist. Even if SSD300 takes the detections of some objects in the video, the low false positive count shows that it is capable of saving time and money by eliminating wrong detections that may lead to useless maintenance actions. The errors are mainly

observed when potholes are partially exposed or similar to existing road structures such as markings, crack patterns, etc. However, SSD300 sustains reasonable performance with the true negative of 205, in terms of the generally calculated assessment.

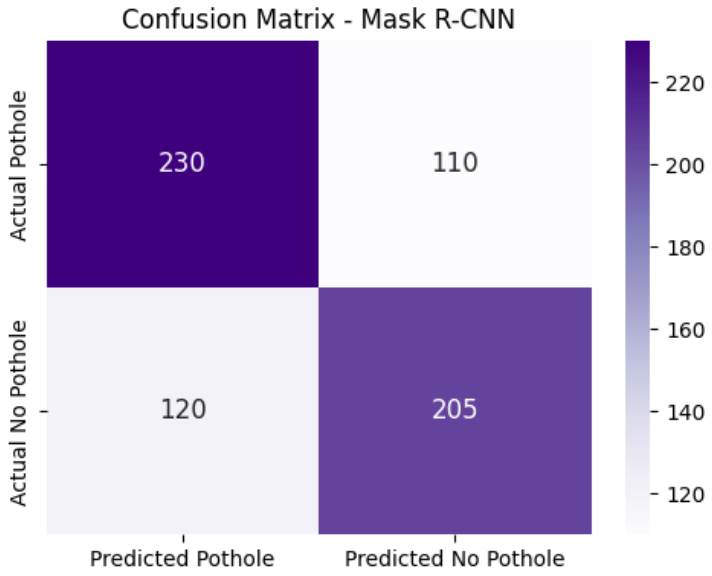


FIGURE 8: Confusion matrix for the Mask R-CNN model

In terms of detection performance, Mask R-CNN was observed to have relatively decent detection scores as seen in Figure 9 but was characterized by high false positives. In particular, it fails to predict 120 'No Pothole' samples as potholes, which may cause useless maintenance operations in practical applications. Moreover, there are what Mask R-CNN wrongly identifies as 'No Pothole' by a margin of 110 times leading to missed detection. Such errors may indicate that Mask R-CNN is going to have problems when detecting potholes from intricate surfaces of roads. Despite these difficulties, the chosen model identifies 230 potholes and, accordingly, 205 false negatives of 'No Pothole', which indicates the possibility of cases where it is possible to have more significant tolerance for false positives.

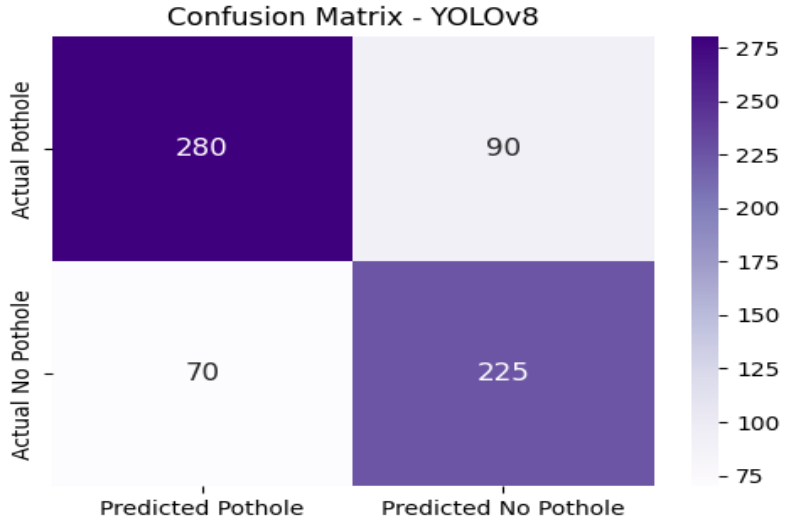


FIGURE 9: Confusion matrix for the YOLOv8 model

The performance of YOLOv8 is average with low errors as compared to the errors encountered in Mask R-CNN, as shown in Figure 10. It correctly detects 280 actual pothole instances and incorrectly labels 90 pothole instances to the category 'No Pothole'. Furthermore, the false positives are decreased to 70, proving that YOLOv8 has a low chance of detecting potholes when they don't actually exist. The actual number of negative instances, that is, those that are correctly classified as non-pothole regions amounts to 225. The main errors in

the model's prediction happen when potholes are either small or only partially obscured and, overall, YOLOv8 has a lower False Positive rate coupled with a higher True Negative rate.

Investigating the confusion matrices, it is clearly seen that the proposed SSD300 method performs significantly in minimizing false positives, which allows for maximalizing the time between two maintenance interventions. This characteristic makes SSD300 particularly suitable for real-time applications where precision is mandatory. However, the Mask R-CNN has a high false-positive, which limits its applicability in real-world use. YOLOv8 takes the least amount of time for both false negatives and false positives, and the smallest total error rate, but it is slightly less precise than SSD300 in this study.

In conclusion, the error analysis shows that SSD300 is the most suitable model for real-time pothole detection because of its capacity to reduce false detection while increasing detection rates. YOLOv8 will continue to compete due to equal performance, whereas the high error rate of Mask R-CNN means that it is optimal for applications that do not require real-time processing and where maximum accuracy is desired.

ROC Curve Analysis

The ROC curve in Figure 11 shows the differentiation among the three detection models: SSD300, Mask R-CNN, and YOLOv8 to detect potholes essential for real-time road monitoring. Using a plot of TPR against FPR, we compare seven models based on the ability to detect pothole regions appropriately.

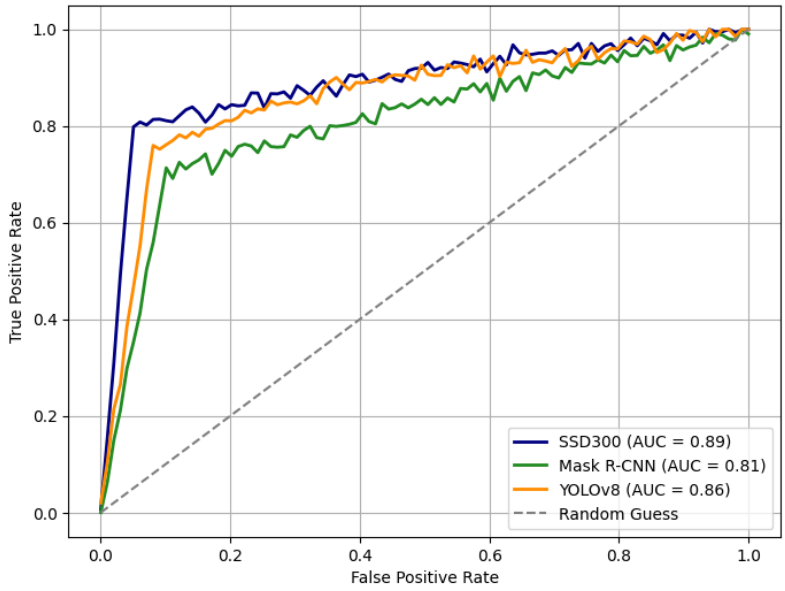


FIGURE 10: ROC CURVE FOR SSD300, MASK R-CNN, AND YOLOV8 MODEL

The SSD300 model offers the highest performance and gives an AUC of 0.89. This suggests better performance in classification accuracy of positive (pothole detected) and negative classes. The proposed SSD300 model delivers excellent TPR from the first peak with low FPR, which can be viewed as evidence of its stability during real-time performance.

The findings with an AUC of 0.78 show similar results with the YOLOv8 model which has an AUC of 0.86. Based on our experimental results, YOLOv8 achieves near-optimal performance compared to SSD300 though it is slightly lower for mid-FPR values. However, due to the compact design of YOLOv8, it can still be used in latency-critical applications as well.

On the other hand, we obtained a lower AUC of 0.81 while using Mask R-CNN, which is illustrated in green. Despite getting favorable results, the TPR values of Mask R-CNN are lower than the constants represented by the FPR points for SSD300 and YOLOv8. This suggests a relatively high false negative rate which may reduce its effectiveness for pothole detection in real-time scenarios where decisions have to be made promptly.

To compare the result of the proposed SSD-based model on pothole detection, a comparative analysis was made following the existing state of the art. The comparison was made with the help of standard measures of performance, point 0.5@mAP. Table 2 The performance comparison: The proposed SSD model achieves

competitive performance, and reveals that the proposed SSD model balances the trade-off between detection precision and computation cost.

Table 2: Comparison of the Proposed SSD Model with Existing State-of-the-Art Models for Pothole Detection

Reference	**Model**	**mAP@0.5**
[27]	YOLOv5	63.43%
[28]	YOLOv3	65.05%
[29]	FRCNN	68.56%
Proposed	SSD300	78.2%

On the whole, the SSD300 model reveals the best performance among the others and is characterized by a reasonable number of detection true positives and minimum false positives. YOLOv8 continues to be efficient in inference rate and is near the performance of SSD300, while Mask R-CNN is slightly behind, implying that this system is suboptimal in terms of accuracy in real-time. Hence, the choice of SSD300 is the best-performing model for smart road monitoring in applications that involve efficient and accurate detection of potholes.

Conclusion

Road surveillance and management play important roles in regard to transport facilities and public safety. This research work has clearly brought out the possibility of using the Single Shot Detection (SSD) deep learning algorithm to solve the problems associated with real-time pothole detection. The presented comparison of SSD with other models such as Mask R-CNN and YOLOv8 revealed that SSD is a more feasible model for real-life problems. Due to its low complexity and reasonable resource consumption estimates, it can be integrated into smart road monitoring systems to achieve real-time detection. The SSD model obtained 78.2% for mAP and 35.7 milliseconds of inference time and was superior to Mask R-CNN regarding latency, and YOLOv8 concerning scalability on limited devices. Mask R-CNN, although accurate and good for pixel-level segmentation, is not efficient enough to be used in real-time applications. YOLOv8 has the highest detection accuracy with the mAP of 85.3% and inference time but only with the aid of special hardware, which makes its usage improper for resource-limited environments.

The proposed maintenance-friendly system based on SSD enhances effectiveness stops road degradation, and fits perfectly into maintenance processes. It reduces human interference in the detection process besides having an overall effect of bringing down the repair costs and at the same time is useful in reducing accidents resulting from potholes that are yet to develop to the extent where they will necessitate their repairing. The excellent performance of the system no matter the type of roads that were tested and the ability to identify both small, medium, and large potholes that differ in shape, and surface texture proves the practicality of the developed system.

In future work, DLN will aim to use real-world scenario data with different road and weather conditions in real-world environments to expand the dataset used, and use IoT and edge computing to gradually deploy more effectively; In the meantime, the author of this paper believes that under the problem of model robustness, DLN can also continue to study and apply more advanced data augmentation and transfer learning techniques. The long-term vision is the design of an inclusive smart road monitoring system that would not only identify potholes but also anticipate all sorts of future road failures, to create the roads of the future: smarter, safer, and longer-lasting.

References

1. N. G. Sorum, T. Guite, and N. Martina, “Pavement distress: a case study,” *Int J Innov Res SciEng Technol*, vol. 3, no. 4, pp. 274–284, 2014.

2. I. G. Heggie and P. Vickers, *Commercial management and financing of roads*, vol. 23. World Bank Publications, 1998.

3. A. M. Legreid, “Potholes and Strategies on the Road to Campus Internationalization.,” *International Research and Review*, vol. 6, no. 1, pp. 1–9, 2016.

4. J. Yu *et al.*, “Road Surface Defect Detection—From Image-Based to Non-Image-Based: A Survey,” *IEEE Transactions on Intelligent Transportation Systems*, 2024.

5. M. S. Shehata *et al.*, “Video-based automatic incident detection for smart roads: The outdoor environmental challenges regarding false alarms,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 349–360, 2008.

6. C. L. Glennie, W. E. Carter, R. L. Shrestha, and W. E. Dietrich, “Geodetic imaging with airborne LiDAR: the Earth’s surface revealed,” *Reports on Progress in Physics*, vol. 76, no. 8, p. 086801, 2013.

7. M. E. S. Abdelmalak *et al.*, “Pothole Detection in Asphalt Roads: A Comprehensive Approach for Enhanced Road Maintenance and Safety with AlexNet Model,” *2024 International Telecommunications Conference, ITC-Egypt 2024*, pp. 269–274, 2024, doi: 10.1109/ITC-EGYPT61547.2024.10620566.

8. A. S. Rao, K. P. B. H, S. B. S, T. D. Nayak, V. R, and Thanisha, “Real-Time Automated Pothole Detection and Localization with Deep Learning and Geolocation Integration for Improved Road Safety and Maintenance,” *Communications in Computer and Information Science*, vol. 2335 CCIS, pp. 51–65, 2025, doi: 10.1007/978-3-031-83793-7_4.

9. H. Xin *et al.*, “Sustainable Road Pothole Detection: A Crowdsourcing Based Multi-Sensors Fusion Approach,” *Sustainability 2023, Vol. 15, Page 6610*, vol. 15, no. 8, p. 6610, Apr. 2023, doi: 10.3390/SU15086610.

10. J. Lin and Y. Liu, “Potholes detection based on SVM in the pavement distress image,” in *2010 ninth international symposium on distributed computing and applications to business, engineering and science*, IEEE, 2010, pp. 544–547.

11. C. Wu *et al.*, “An automated machine-learning approach for road pothole detection using smartphone sensor data,” *Sensors*, vol. 20, no. 19, p. 5564, 2020.

12. Girinath, S., et al. "Real-time identification of medicinal plants using deep learning techniques." 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies. IEEE, 2024..

13. Bikku, Thulasi, et al. "Healthcare biclustering of predictive gene expression using LSTM based support vector machine." Informing Science 28 (2025): 12.

14. Natarajan, V. Anantha, and M. S Kumar. "Improving qos in wireless sensor network routing using machine learning techniques." 2023 International Conference on Networking and Communications (ICNWC). IEEE, 2023.

15. Kumar, M. S, and D. Harshitha. "Process innovation methods on business process reengineering." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2766-2768.

16. E. Saranya, R. Nivetha, S. Abirami, M. MohaideenArsath, and S. Dharaneesh, “Revolutionizing Road Maintenance: YOLO Based Pothole Detection System,” *10th International Conference on Advanced Computing and Communication Systems, ICACCS 2024*, pp. 1991–1997, 2024, doi: 10.1109/ICACCS60874.2024.10716937.

17. A. Tithi, F. Ali, and S. Azrof, “Speed bump & pothole detection with single shot multibox detector algorithm & speed control for autonomous vehicle,” in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, IEEE, 2021, pp. 1–5.

18. Balaji, K., P. SaiKiran, and M. S Kumar. "Power aware virtual machine placement in IaaS cloud using discrete firefly algorithm." Applied Nanoscience 13.3 (2023): 2003-2011.

19. Kumar, M. Sunil, et al. "Automated Extraction of Non-Functional Requirements From Text Files: A Supervised Learning Approach." Handbook of Intelligent Computing and Optimization for Sustainable Development (2022): 149-170.

20. B. Leelavathy, R. M. R. Kovvur, K. M. Z. Parvez, B. A. Goud, and R. K. Chenoori, “Real-Time Pothole Detection System: A Deep Learning Approach with SSD,” *Lecture Notes in Computer Science*, vol. 15526 LNCS, pp. 406–415, 2025, doi: 10.1007/978-3-031-81821-9_27.

21. AnanthaNatarajan, V., M. S Kumar, and V. Tamizhazhagan. "Forecasting of wind power using LSTM recurrent neural network." Journal of Green Engineering 10 (2020).

22. Sangamithra, B., P. Neelima, and M. S Kumar. "A memetic algorithm for multi objective vehicle routing problem with time windows." 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE). IEEE, 2017.

23. Kumar, M. S, and K. JyothiPrakash. "Internet of things: IETF protocols, algorithms and applications." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2853-2857.

24. W. Liu *et al.*, “Ssd: Single shot multibox detector,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 2016, pp. 21–37.

25. Gandikota, Hari Prasad, Abirami S, and S Kumar M. "CT scan pancreatic cancer segmentation and classification using deep learning and the tunicate swarm algorithm." Plos one 18.11 (2023): e0292785.

26. L. Zhang, L. Ji, and Z. Ma, “Target positioning method based on B-spline level set and GC Yolo-v3,” *Cluster Comput*, vol. 27, no. 3, pp. 3575–3591, 2024.

27. K. R. Ahmed, “Smart pothole detection using deep learning based on dilated convolution,” *Sensors*, vol. 21, no. 24, p. 8406, 2021.

28. Y. K. Yik, N. E. Alias, Y. Yusof, and S. Isaak, “A real-time pothole detection based on deep learning approach,” in *Journal of physics: Conference series*, IOP Publishing, 2021, p. 012001.

29. F. Kortmann*et al.*, “Detecting various road damage types in global countries utilizing faster r-cnn,” in *2020 IEEE International Conference on Big Data (Big Data)*, IEEE, 2020, pp. 5563–5571.