INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 5 (2025) 402-421
https://publisher.uthm.edu.my/ojs/index.php/ijie

Pothole Detection and Classification Using Enhanced Efficient-Net Optimized by Advanced Manta-Ray Algorithm

Muhammad Numan Ali Khan[1]*, Mohd. Norzali Haji Mohd[1], Tasiransurini Ab Rahman[1], Nuzhat Khan[2], Muhammed Paend Bakht[3]

[1] *Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400, Johor, MALAYSIA*
[2] *Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, MALAYSIA*
[3] *Department of Electrical Engineering, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta 87300, PAKISTAN*

*Corresponding Author: he210013@student.uthm.edu.my
DOI: https://doi.org/10.30880/ijie.2025.17.05.031

Article Info

Received: 7 May 2025
Accepted: 26 August 2025
Available online: 30 August 2025

Keywords

Pothole detection, efficient-net, manta ray optimization, image classification, machine learning

Abstract

Object detection powered by neural networks has transformed artificial intelligence applications with notable advancements in numerous domains, including the automated detection of road potholes. However, existing deep learning models that rely on conventional optimizers such as Adam face challenges like slow convergence and susceptibility to local minima. These limitations often result in suboptimal performance in complex real-world scenarios. To address these challenges, this study introduces a novel methodology to improve pothole detection and classification accuracy and efficiency by integrating the Manta Ray Foraging Optimization (MRFO) algorithm into Efficient-Net. The MRFO, inspired by the collective foraging behavior observed in manta rays, is implemented as a replacement for the conventional Adam optimizer (AO) in Efficient-Net. This integration strengthens feature extraction by managing the trade-off between global exploration and local exploitation. Using a comprehensive pothole dataset sourced from Kaggle, the proposed model successfully overcomes prominent optimization challenges, leading to a substantial gain in classification accuracy from 84% to 93%. Comprehensive experiments were conducted across the B0, B1, and B2 configurations of Efficient-Net, comparing the performance of MRFO against traditional optimization methods. Results consistently demonstrate that MRFO significantly enhances pothole detection and classification capabilities. This study highlights the potential of MRFO as a robust optimization tool for real-world object detection tasks that can further improve its broader applications in intelligent transportation and beyond.

1. Introduction

The rapid expansion of urban infrastructure and transportation systems has emphasized the need for automated solutions to maintain road quality and safety [1]. Potholes, one of the most prevalent road defects, not only compromise driving conditions but also pose significant risks to vehicle safety and passenger comfort [2]. Effective detection and classification of potholes are crucial for timely maintenance and cost-effective resource allocation. Traditional methods of pothole detection relying on human inspectors are increasingly unsustainable due to their inherent limitations. Manual inspection is time-intensive and leads to delays in detecting potholes. Consequently, road defects that are not addressed in a timely manner cause further deterioration. The process is also financially burdensome, as inspection vehicles are required to traverse damaged roads frequently. Ultimately, this increases maintenance costs and fuel consumption and reduces operational lifespans. Furthermore, manual inspections expose personnel to significant safety risks, particularly on busy highways or in adverse weather conditions [3]. The subjective nature of human evaluation can also result in inconsistencies, with defects being misclassified or overlooked entirely. Additionally, this approach is not easily scalable to accommodate the growing need for consistent, real-time monitoring of extensive road networks. In contrast, AI-based automated detection provides a more sustainable solution using advanced algorithms for accurate, efficient, and scalable pothole identification [4]. AI-assisted pothole detection will likely eliminate many of the drawbacks of manual inspection methods. In addition, traditional methods for pothole detection, such as manual inspection and conventional image processing techniques, are often labor-intensive, prone to errors, and inefficient when applied at large scales.

In recent years, artificial intelligence (AI) and deep learning advancements have demonstrated immense potential to address these challenges with automated pothole detection through object detection frameworks. Among various AI techniques, convolutional neural networks (CNNs) have emerged as a powerful feature extraction and classification tool [5]. Architectures like Efficient-Net have gained popularity for their ability to optimize model accuracy while minimizing computational complexity [6]. However, the performance of these models heavily relies on the choice of optimization algorithms, which play a critical role in minimizing loss functions and improving convergence during training [7]. The Adam optimizer (AO), a popular choice in machine learning, is widely used in deep learning due to its adaptive learning rate and efficient handling of sparse gradients. However, it has certain drawbacks that can limit its performance in complex optimization tasks [8]. AO is efficient and widely used in CNN optimization, but it can converge too quickly to suboptimal solutions because of its aggressive learning rates and limited search. It is also prone to poor generalization, overfitting, and sensitivity to hyperparameters, which can lead to instability and inconsistent performance. These limitations make AO less suitable for tasks that demand a comprehensive search of the solution space while maintaining precision in fine-tuning model parameters, such as deep learning-based pothole detection.

To balance global exploration and local exploitation Manta Ray Foraging Optimization (MRFO) alternatively provides a bio-inspired approach to solve optimization problems. It is currently one of the most widely used algorithms for hyperparameter optimization. The MRFO algorithm draws inspiration from the way manta rays and mimics their natural foraging patterns of navigating ocean waters in search of plankton. Unlike AO, MRFO explores the search space more comprehensively [9]. It mimics the collaborative and strategic foraging behavior of manta rays to avoid premature convergence. Its ability to dynamically adapt to optimization challenges makes it a robust alternative for complex tasks like feature extraction and classification in deep learning. By addressing the limitations of AO, MRFO can provide a pathway to achieve improved convergence, better generalization, and enhanced performance in tasks such as pothole detection [10].

In this study, we propose a novel integration of the MRFO algorithm with the Efficient-Net architecture to improve pothole detection and classification accuracy. By replacing the traditional AO with MRFO, our approach aims to address key optimization challenges associated with the conventional optimizer. This work contributes to the growing body of research on applying bio-inspired optimization techniques to deep learning. The proposed framework provides a solution for intelligent real-time road infrastructure monitoring and maintenance. This paper is organized as follows: Section 2 reviews related works on pothole detection, optimization algorithms, and Efficient-Net architectures. Section 3 outlines the proposed methodology, detailing the integration of MRFO with Efficient-Net. Section 4 presents the experimental setup and results, followed by an analysis of the findings. Finally, Section 5 concludes the study and highlights future research directions.

1.1 Related Work

Traditional methods for pothole detection rely on image processing techniques, including edge detection and texture analysis [11]. Recent advancements in deep learning have made it possible to use convolutional neural networks (CNNs) for more accurate road defect detection [12]. However, these models often require significant computational resources [13]. The application of CNNs for pothole detection has gained substantial attention in recent years. Studies have demonstrated the suitability of CNNs in identifying potholes from road images because of their capacity to extract features from complex datasets automatically [5]. Techniques such as transfer learning, where pre-trained CNN architectures are fine-tuned for specific datasets, have also been employed to improve

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421

detection accuracy [12]. As AI-assisted pothole detection is a complex, multifaceted task that requires machine vision, high-quality image datasets, comprehensive image preprocessing, appropriate feature extraction, and strong algorithms with properly tuned hyperparameters.

The existing literature revolves around the approaches to pothole detection, the advancements in neural network architectures, and the optimization methodologies employed. Numerous studies have explored the application of deep learning architectures in detecting and classifying potholes. For instance, the efficiency of a basic CNN architecture for pothole detection using image datasets, achieving commendable results in a controlled environment, is demonstrated in [14]. While various configurations of CNN are compared with object detection tools, it was found that CNN is a computationally expensive model [5]. The limited depth and feature extraction capabilities of traditional CNNs hinder their performance in complex real-world scenarios involving varying light conditions and obstructed views of potholes. In response to these limitations, researchers have turned to advanced architectures such as Efficient-Net, which optimizes performance through a compound scaling method that simultaneously adjusts depth, width, and resolution [15].

Several studies have reported improved accuracy in image classification tasks using Efficient-Net compared to standard CNNs [16]. The study [17] proposed a Federated Unified Network Aggregation (FedUNA) algorithm for pothole classification using the Efficient-Net architecture within a Federated Learning (FL) framework. To assist the impaired in pothole detection, [6] proposed a YOLO algorithm that achieved 82.7% accuracy. Similarly, [18] preferred the Particle Swarm Optimization (PSO) algorithm integrated with the Conventional Forest model for pothole detection. Meanwhile, [19] proposed Efficient-Net with AO for road defect detection. Efficient-Net, a family of CNN architectures, achieved state-of-the-art performance using a compound scaling method for multiclass classification with fewer parameters. Its ability to balance model depth, width, and resolution proved it suitable for resource-constrained environments. However, the experiment preferred a conventional Efficient-Net with the gradient-based algorithm AO. In highly complex problems, gradient-based optimization algorithms like PSO and AO stagnate into premature convergence to local minima, leading to suboptimal outcomes. The existing studies have not adequately addressed these limitations, leaving room for further innovation.

Despite these advancements, CNN-based approaches often suffer from limitations such as overfitting on small datasets [5], sensitivity to variations in lighting and environmental conditions [20], and high computational demands, which restrict their scalability and real-time applicability. Despite the importance and widespread adaptation of pothole detection models, existing studies fail to comprehensively tackle challenges related to the complexity, large number of parameters, accuracy, and real-time performance of classification models. Previously proposed works do not adequately address the resource limitations, leaving room for further innovation. In a resource-constrained environment, Efficient-Net, which is from a family of CNN architectures, uses a compound scaling method to achieve state-of-the-art performance with fewer parameters. Its unique features include a balanced depth, width, and resolution in the model [21]. However, even with the strengths of Efficient-Net, traditional optimization methods like AO and stochastic gradient descent (SGD) often struggle to escape local minima in complex optimization spaces. It limits the ability of AO and SGD to fine-tune models effectively. Metaheuristic algorithms, such as PSO and genetic algorithm (GA), have also been explored to enhance CNN performance and overcome these limitations [22]. While these methods demonstrate potential, they come with constraints, including convergence to suboptimal solutions and dependence on extensive parameter tuning.

In this context, the MRFO algorithm emerges as a promising alternative. Inspired by the foraging behavior of manta rays, MRFO offers a robust global optimization strategy to balance exploration and exploitation, making it well-suited for optimizing Efficient-Net in terms of hyperparameters and feature selection. By addressing the limitations of existing methods, MRFO presents a novel pathway toward developing sustainable, efficient, and accurate deep learning models for pothole detection. Abdul Razak and his team [6] proposed an improved version of the MRFO algorithm that uses a sorting technique based on Pareto's game theory to rapidly construct a high-quality Pareto front. This enhanced algorithm was named NSMRFO. It was tested on various standard benchmark functions, and its performance was analyzed using hypervolume-based statistical metrics, showing comparisons with other optimization methods. Additionally, a hybrid optimization approach was applied by combining characteristics of two metaheuristic techniques to address the drawbacks of individual algorithms. MRFO was benchmarked against other advanced optimizers, standard test functions, and real-world engineering design problems. The comparison results indicated that MRFO consistently outperformed other algorithms. Moreover, in practical engineering applications, the algorithm proved effective in terms of speed, accuracy, and computational efficiency when handling complex optimization tasks [1]. This study, however, did not consider the pothole detection problem. Despite the progress made, the potential benefits of MRFO in image processing tasks remain largely unexplored. Similarly, the nature-inspired method MRFO for pothole classification has not been thoroughly investigated in the existing literature. Importantly, there is a clear and unaddressed gap regarding the synergistic application of enhanced Efficient-Net combined with MRFO for accurate pothole detection and classification.

This study aims to bridge this gap, proposing a novel framework, EFMANTA, that combines the strengths of Efficient-Net and MRFO for pothole detection. Thus, the current research contributes to the growing knowledge of smart infrastructure monitoring for timely maintenance to overcome urban transportation challenges.

2. Materials and Methods

Deep learning for pothole detection typically involves several key steps: data collection, labeling, model design, training, evaluation, and deployment. A substantial dataset comprising both images of potholes and non-pothole areas is collected to facilitate the training and testing of the deep learning model. The step-by-step process of adopted methods and a description of the utilized datasets are provided in the subsequent sections.

2.1. Pothole Detection Dataset

Data quality and quantity are paramount in image processing and machine vision tasks. For this experiment, a dataset is sourced from Kaggle (https://www.kaggle.com/datasets/atulyakumar98/pothole-detection-dataset). It is specifically curated for the pothole detection problem [23]. The dataset initially contained more than 600 images representing 352 normal road conditions and 329 potholes. Each image was annotated with bounding boxes to mark the region of interest (ROI) pothole and class labels to distinguish between pothole and non-pothole areas. To enhance its quality and size, the initially obtained raw dataset is transformed during preprocessing. The description of the raw and transformed dataset is given in Table 1. The dataset contains significant illumination variations and occlusion challenges that reflect real-world conditions. Images were captured under various lighting scenarios, including daylight, shadowed areas, and low-light environments. Additionally, the dataset incorporates several occlusion scenarios such as vehicle shadows, water puddles, and debris coverage, which introduce authentic real-world detection challenges.

Table 1 *Pothole detection dataset*

Attribute	Description
Dataset Title	Pothole Detection Dataset
Total Images	Originally 681images 352 normal 329 and potholes
Annotation	Each image containing a pothole is properly annotated
Data Augmentation Output	Expanded to 2043 images through augmentation techniques 1056 normal and 987 potholes
Primary Objective	Detect and classify potholes in road images
Class Labels	Two categories: Normal Road and Pothole
Augmentation Techniques	Includes transformations such as rotation, scaling, flipping, pixel inversion, etc.
Dataset	To increase variability in training data and boost model generalization

In computer vision-based approaches, potholes are detected based on different visual characteristics such as circular or irregular shapes, darker appearance in contrast to the surrounding surface, and rough texture compared to smoother surrounding areas [24]. These attributes are utilized to highlight the boundaries of potholes within an image. Feature extraction models assess the texture, shape, and dimensions of the damaged region, identifying defects by analyzing and contrasting these features with those of the adjacent surface [25]. For better understanding, sample normal images and pothole images with bounding boxes from the dataset are provided in Figure 1.

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421

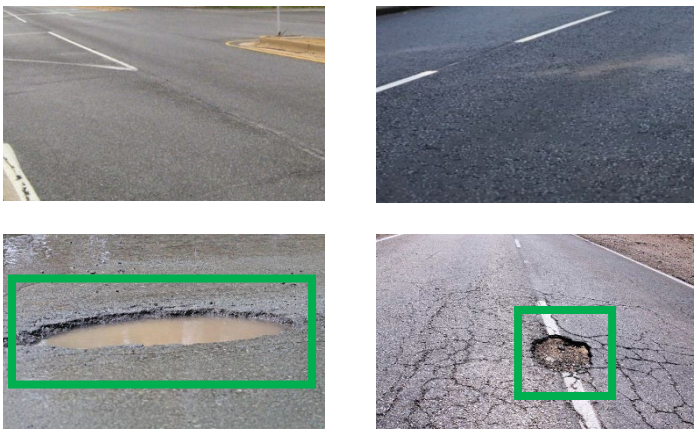


Fig. 1 *Sample images from the pothole detection dataset*

2.2. Data Preprocessing

Preprocessing steps included resizing images to a uniform resolution of 224x224x3, grayscale conversion for reducing computational complexity, and feature scaling to normalize pixel values. Data augmentation techniques, including rotation, flipping, and contrast adjustments, were applied to obtain sufficient training samples for the proposed model. Images were rotated at angles such as 60°, 90°, and 180°. During this transformation, the new images were generated from rotated images. Data augmentation not only enhanced the diversity of the training dataset, increasing the number of samples to 2000, but it also contributed to improved generalization and robustness of the classification model. Flipping changed the direction of the object while zooming in, which made ROI extraction convenient for the model. Finally, resized images were transformed into a binary format using One-hot encoding [26]. An overview of the preprocessing pipeline is provided in Figure 2.

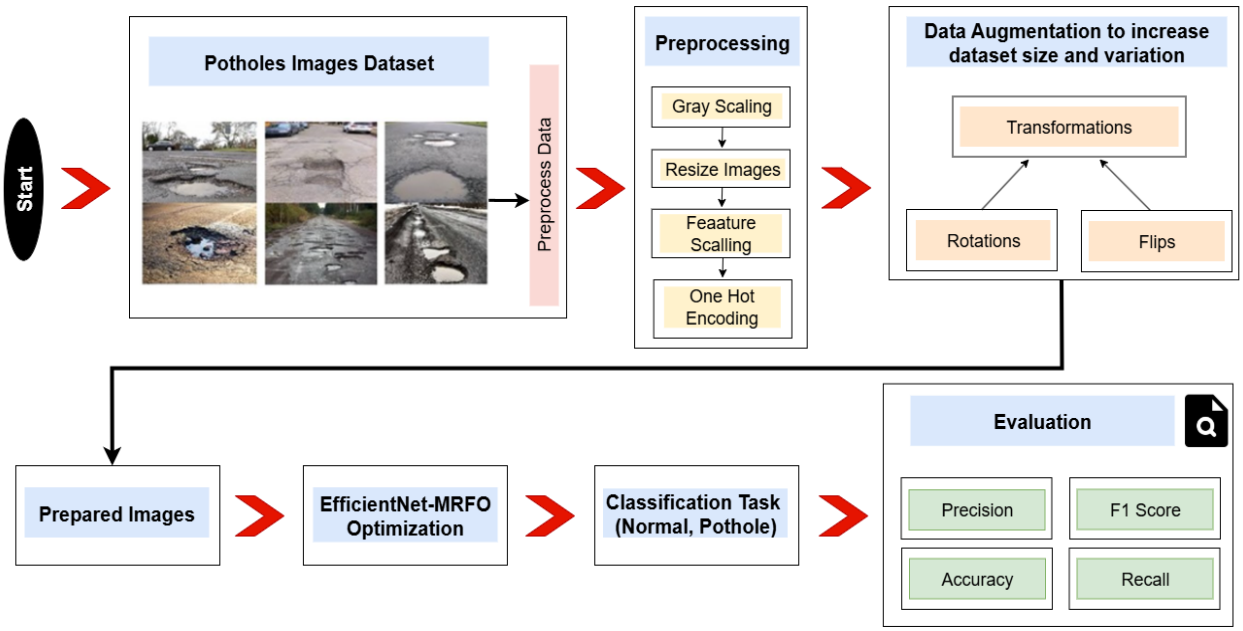


Fig. 2 *Block diagram of the proposed methodology*

2.3. Manta Ray Foraging Optimizer (MRFO)

This section introduces the working principles and provides insight into the proposed optimization algorithm, MRFO. The discussion aims to interpret how MRFO enhances the performance of Efficient-Net by optimizing hyperparameters and selecting the most relevant features. The algorithm strategically balances exploration and exploitation through three key mechanisms: chain foraging, cyclone foraging, and somersault foraging [27]. The mechanisms are briefly described in subsequent sections.

2.3.1 Chain Forging

Manta rays engage in a behavior known as chain foraging in which they align themselves in a head-to-tail formation and move collectively to maximize their capture of plankton. Typically, more than 50 individuals form these synchronized chains to improve feeding efficiency. In MRFO, this collective search strategy is modeled mathematically by updating each manta ray's position based on the leading solution in the population. The position update rule for each individual is defined as follows:

$$x_i^{t+1} = \begin{cases} x_{best}^t + r \,.\, [x_{best}^t - x_i^t] + \alpha \,.\, [x_{best}^t - x_i^t], i = 1 \\ x_{best}^t + r \,.\, [x_{i-1}^t - x_i^t] + \alpha \,.\, [x_{best}^t - x_i^t], i = 2, \ldots . N \end{cases} \quad (1)$$

$$\alpha = 2 \,.\, r \,.\, \sqrt{|\log(r)|} \quad (2)$$

x^t is the manta ray position at iteration t, r is a random vector, x_{best}^t indicates the height concentration of plankton, α is the weight coefficient, and N is the number of manta rays.

2.3.2 Cyclone Forging

When plankton concentration greatly exceeds the number of foraging manta rays, the group tends to form a vortex-like arrangement. In this formation, individuals swim in spiral chains aimed at the target. Unlike WOA, MRFO incorporates a head-to-tail dependency within this vortex behavior. The position update equation for cyclone foraging is given as:

$$x_i^{t+1} = \begin{cases} x_{best}^t + r \,.\, [x_{best}^t - x_i^t] + \beta \,.\, [x_{best}^t - x_i^t], i = 1 \\ x_{best}^t + r \,.\, [x_{i-1}^t - x_i^t] + \beta \,.\, [x_{best}^t - x_i^t], i = 2, \ldots . N \end{cases} \quad (3)$$

Here β is the weighting coefficient.

2.3.3 Somersault Forging

Manta rays demonstrate a unique foraging behavior known as somersaulting. During this behavior, they swim around areas with plankton and execute backward somersaults in a random, circular pattern. Within the framework of MRFO, plankton serves as a focal point, and manta rays constantly modify their positions around the best solution identified to date. This optimization technique is aimed at maximizing their food intake. The mathematical formula that represents the somersault foraging process is as follows:

$$x_i^{t+1} = x_i^t + S\,(r_2 \cdot x_{best}^t - r_3 \cdot x_i^t) \quad (4)$$

S indicates the somersault weight determining the range of somersault behaviors, where $S = 2$; r_2 and r_3 are two random vectors within [0, 1] x_i^t is the velocity of manta ray i at time t . x_{best}^t is the fitness of the best manta ray at time t x_i^t is the fitness of manta ray i at time t in image dataset MRFO feature map updating their position w.r.t feature:

$$x_i^t = (x_i^t \,.\, r1) - (x_{i+1}^t \,.\, r2) i = 1,2,3 \ldots ., \quad (5)$$

In this equation x_i^t,starting position of manta ray with respect to t is a random vector x_{i+1}^t is the best position of manta ray

$$Ntotal\ Agent = (x_i^t \cdot r_1) \cdot (x_{best}^t - x_i) + \alpha\,(x_i + 1\,x_1) = 1,2,3, \ldots . , N \quad (6)$$

x_i^t starting position of manta ray with respect to t is the random vector x_{i+1}^t is best position of a manta ray. N_{total} Agent α = weight of the coefficient.

2.4. MRFO for Optimization

To apply MRFO for optimization problems, the search space boundaries are defined starting from the upper bound (first feature) and ending with the lower bound (last feature). Each initial feature is computed based on time-step

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421

data, with adjustments involving random vectors and weighted coefficients. This process iteratively continues until a stopping criterion is met, typically when the output value falls below a set threshold, which was set to 0.5 in this experiment. This approach ensures consistent feature optimization during the search process.

2.5. Efficient-Net B0

Efficient-Net B0 is a CNN architecture designed to achieve optimal performance through a novel technique known as compound scaling. Traditional CNN architecture often scales the model by arbitrarily increasing depth (number of layers), width (number of channels per layer), or image resolution, leading to inefficient use of computational resources and unbalanced models. Efficient-Net B0, by contrast, employs a principled scaling method that simultaneously and uniformly adjusts the network for depth (d), width (w), and input resolution (r) using a compound coefficient denoted as ϕ. This results in a well-balanced architecture that achieves better accuracy and efficiency than previous models with a comparable number of parameters. Compound scaling is defined mathematically as:

$$d = \lambda^{\phi}, w=\mu^{\phi}, r= \nu^{\phi} \quad (7)$$

Here λ, μ, and ν are the constants controlling the rate of scaling for each dimension. These constants are determined via a grid search on the baseline model and satisfy the constraint:

$$\lambda \cdot \mu^2 \cdot \nu^2 \approx 2 \quad (8)$$

This compound scaling approach of Efficient-Net B0 in larger images creates balanced adjustments across all dimensions. The network then expands its receptive field and captures fine-grained details. By integrating depth, width, and resolution into a uniform scaling framework, Efficient-Net B0 delivers state-of-the-art performance with improved computational efficiency. For tasks involving high-resolution images or where detailed spatial information is essential, such as in surface defect detection, this balanced approach significantly enhances performance. A conceptual layout of the underlying Efficient-Net B0 architecture is given in Figure 3.

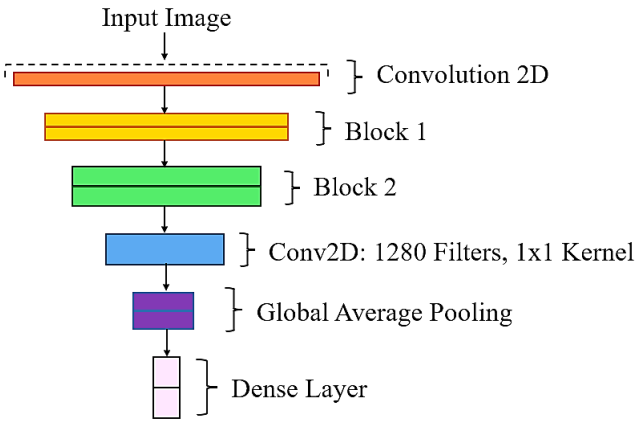


Fig. 3 *Schematic diagram of the architecture of the basic Efficient-Net B0 variant*

Efficient-Net B0 is the baseline model in the Efficient-Net family, from which other variants like Efficient-Net B1 and Efficient-Net B2 are derived. These advanced versions employ higher values of the compound coefficient ϕ, resulting in progressively deeper, wider networks capable of processing higher-resolution images. As a result, Efficient-Net B1 and B2 exhibit enhanced capacity to capture complex patterns in data and perform better on large-scale image classification tasks. However, these gains come at the cost of increased computational requirements.

Despite the capabilities of B1 and B2, Efficient-Net B0 is particularly well-suited for real-world applications like pothole detection, where computational efficiency, inference speed, and model size are critical factors, especially when deploying on resource-constrained platforms such as mobile devices or embedded systems in vehicles. Pothole detection often involves analyzing street-level imagery captured under varying conditions. Therefore, the model must be accurate and fast to enable real-time detection and response. Efficient-Net B0 achieves this balance effectively, offering a robust solution without overwhelming hardware limitations.

In this research, the Efficient-Net B0 architecture was adopted and customized for pothole detection. The architecture begins with a standard 3×3 convolutional stem layer, followed by a series of MBConv (Mobile Inverted Bottleneck Convolution) blocks. These blocks feature depth-wise separable convolutions, expansion layers, squeeze-and-excitation modules, and skip connections. The MBConv blocks are organized in stages, with

varying kernel sizes (mostly 3×3 and 5×5), expansion factors, and strides to downsample and extract hierarchical features. After these blocks, a 1×1 convolution layer aggregates features, followed by a global average pooling layer and a final fully connected (dense) layer for classification output as shown in Figure 4. The full architecture of Efficient-Net B0 is organized into several stages. Each stage contains a series of MBConv blocks with specified configurations. With this architectural design, Efficient-Net B0 efficiently captures low-level textures (such as cracks and edges) and high-level structural patterns (including the shape and context of potholes) in road surfaces.

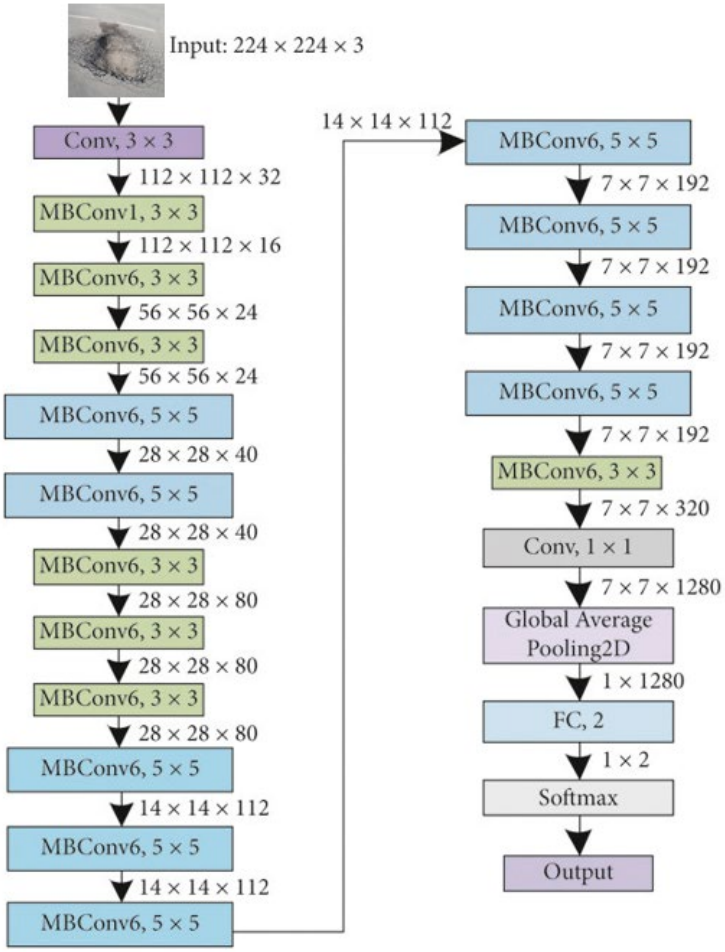


Fig. 4 *Detailed architecture of the Efficient-Net B0 model used in this research*

A key innovation in Efficient-Net B0 is the use of MBConv (Mobile Inverted Bottleneck Convolution) blocks. These blocks are derived from MobileNetV2 and are designed for computational efficiency. Each MBConv block consists of:

- Expansion layer with a 1×1 convolution increases channels to a higher-dimensional feature space.
- Depth-wise convolution applies a separate spatial filter to each input channel independently to significantly reduce computation while preserving spatial structure.
- Squeeze-and-Excitation (SE) module that adaptively refines channel-wise feature responses using global information to amplify the most informative features.
- Projection layer using a 1×1 convolution to reduce the number of channels back to the original size, allowing for efficient feature compression.
- A skip connection that integrates the input directly with the output when their dimensions align, for efficient gradient propagation by minimizing information degradation between layers.

2.6. Integrated Architecture of MRFO and Efficient-Net:

In this experiment standard Efficient-Net (B0, B1, B2) backbone is utilized as the baseline binary classification model for pothole detection. However, the architecture of the baseline model was modified by integrating MRFO for optimization to overcome the limitations of gradient-based optimizers such as AO. MRFO optimized feature

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421

vectors to ensure that the model focused on the most relevant and discriminative features. Efficient-Net extracted features systematically, beginning with low-level patterns in the initial layers and advancing to more complex and abstract representations in the deeper layers. This process generated feature maps that captured both spatial and contextual information essential for accurate detection. The feature maps were initially compressed into one-dimensional feature vectors, which represented the high-dimensional attributes of each image.

The MRFO algorithm began with a population of randomly generated solutions, each representing a unique subset of features extracted by Efficient-Net. A fitness function based on classification accuracy was applied to evaluate these solutions to identify promising feature subsets. The algorithm iteratively refined the feature selection by dynamically updating the fitness scores of the solutions with a balance between exploration and exploitation of the feature space. This iterative process converged toward an optimal solution by focusing on the most significant feature subsets. The optimized feature maps were then processed through the fully connected layer, where the filter weights were fine-tuned to consolidate the selected features further. This stage incorporated the chain foraging behaviour of MRFO for efficient exploration and exploitation of the feature space. MRFO enhanced the performance of Efficient-Net by optimizing its hyperparameters and feature selection. The process began with an initialization phase, where each manta's position in the solution space was randomly generated, representing a potential combination of hyperparameters or feature subsets. The fitness value for each manta was calculated using a fitness function, which evaluated the classification performance of Efficient-Net under the configuration considered.

In the optimization phase, the algorithm employed a decision mechanism to update the positions of the mantas. Based on a randomly generated variable r, the algorithm alternated between the chain foraging equation and the cyclone foraging equation to adjust the manta's position. If r was less than 0.5, the chain foraging strategy was applied to prioritize exploration. Conversely, if r was 0.5 or greater, the cyclone foraging strategy was used to exploit promising regions of the search space. This dynamic decision-making process provided a balance between exploration and exploitation that is essential for finding optimal solutions. The proposed integrated architecture of MRFO and Efficient-Net is presented in Figure 5.

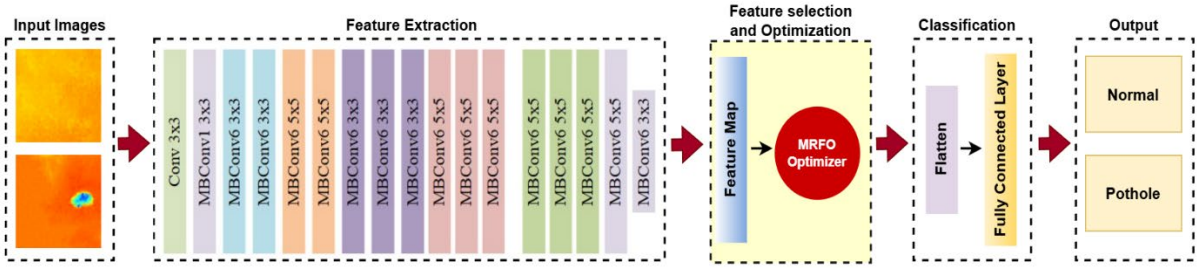


Fig. 5 *Proposed hybrid MRFO-Efficient-Net architecture*

2.3.4 MRFO Feature Optimization Strategy

Efficient-Net processed input images to extract feature maps representing spatial and semantic information, such as edges, contours, and surface variations. These feature maps were then passed to the MRFO algorithm for optimization. MRFO defined a search space that included the extracted feature maps. Afterward, it employs an optimization mechanism to identify the most relevant features. This involved filtering out noisy or redundant data while retaining features critical for pothole segmentation. The MRFO optimization process was guided by a fitness function designed to maximize segmentation accuracy. Once the feature maps are optimized, they are converted into segmentation masks delineating pothole boundaries. The segmentation masks on images were generated during training using the optimized feature maps. These masks were saved in the training dataset. During testing, the optimized segmentation masks were used to segment new images by matching their feature maps to the most similar patterns from the training set. The boundary of a pothole is illustrated in Figure 6 with a segmentation mask generated using MRFO.

Fig. 6 *Pothole labelling*

2.3.5 Iterative Refinement and Feature Selection by MRFO

The iterative nature of MRFO, combined with its dynamic exploration and exploitation strategies, made it an efficient choice for this application. The iterative process of obtaining the highest fitness value continued until the iteration number reached the predefined maximum number of iterations (T_{max}). During each iteration, the algorithm evaluated the fitness function for the updated manta positions and retained the highest fitness value. Once T_{max} was reached, the somersault foraging equation was applied to refine the positions of the mantas further to achieve convergence to the optimal solution. This equation simulated a final exploitation phase by intensively searching for the best solutions to extract the most significant features. The final output of the MRFO algorithm was the optimal feature set derived from the position of the manta ray with the highest fitness value. By integrating MRFO into Efficient-Net, the optimization process effectively enhanced feature selection and hyperparameter tuning, as shown in Fig. 7.

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421

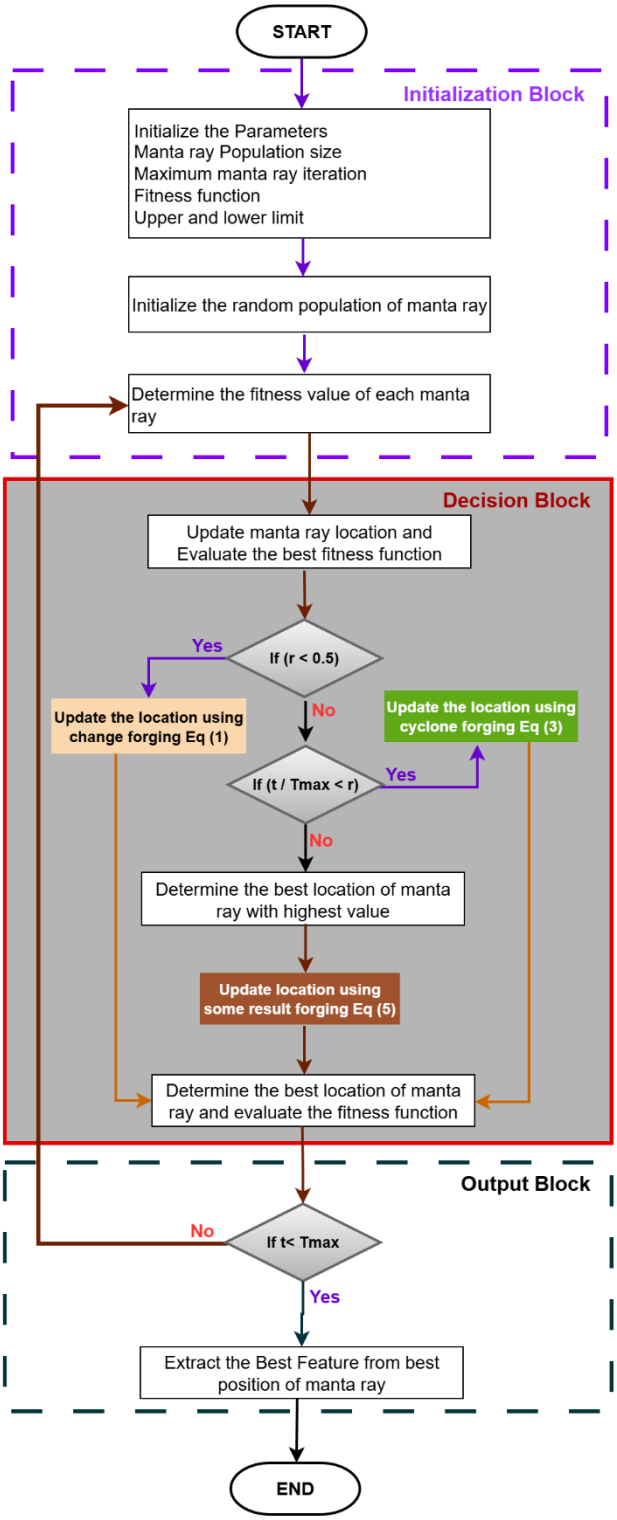


Fig. 7 *Flowchart of the proposed model*

2.3.6 Proposed MRFO Pseudocode

The MRFO algorithm begins with the initialization of a population of agents where each agent represents a candidate solution within the defined search space. The three main operators, including Chain Foraging, Cyclone Foraging, and Somersault Foraging are applied iteratively to update the position of each agent. At each iteration, candidate solutions are guided toward the best-known solution, ensuring a balance between exploration (global search) and exploitation (local refinement). This mechanism enhances convergence compared to traditional optimizers and improves the classification performance of the deep learning model by achieving greater accuracy

and robustness in pothole detection under varying environmental conditions. The pseudocode of the MRFO algorithm is outlined below:

Initialize_MRFO (position, velocity, optimal solution)
Initialize position, velocity, and optimal solution.
Update_MRFO (position, velocity, optimal solution)
Compute the distance between the current manta ray position and the optimal solution.
Adjust the velocity according to this distance.
Modify the position using the updated velocity.
Get_Current_Position(position):
Present the updated position of the manta ray.
Setup_MRFO(population_size, search_bounds, optimal_solution):
Define the number of agents (population_size), search space limits, and initial optimal_solution.
Randomly generate initial positions and velocities for all manta rays within the defined bounds.
Run_MRFO_Optimization(population, max_iterations):
Repeat for the specified number of iterations:
For each manta ray in the population:
Update position using Update_MRFO.
Identify the individual with the best fitness score.
Replace the optimal solution with the best individual position of the manta ray if improved.

2.7. Performance Evaluation, Comparison, and Benchmarking

The performance evaluation of the proposed integrated architecture, which combines Efficient-Net with MRFO, was conducted systematically to ensure its reliability for real-world pothole detection applications. The model was trained using three variations of Efficient-Net: the baseline network B0, a moderately complex network B1, and a larger variant B2, which has increased resolution, depth, width, and complexity [28]. The primary difference between Efficient-Net B0, B1, and B2 lies in their scaling factors for network depth, width, and input resolution, which determine their complexity and performance. These models are part of the Efficient-Net family, where each version incrementally scales up the base architecture (Efficient-Net B0) using the compound scaling method [29]. Each configuration was assessed using standard evaluation metrics, including accuracy, Area under the curve (AUC), Precision, Recall, F1-score, and confusion matrix. Accuracy provided an overall measure of correct predictions, while the confusion matrix detailed the performance in terms of true positives, true negatives, false positives, and false negatives. The outcomes in the form of true positives, true negatives, false positives, and false negatives are also included in the assessment to judge the robustness of the architecture. The performance comparison among the Efficient-Net variants aimed to select the most effective model configuration.

To further highlight the improvements of the proposed network over existing optimization techniques, it was compared against conventional Efficient-Net B0, B1, and B2 models integrated with AO. Besides, additional validation and benchmarking were performed to prove the superiority of the proposed architecture over the standard CNN architecture. Furthermore, the outcomes were thoroughly validated, and the accuracy of the proposed architecture was compared with the existing studies. A step-by-step comprehensive evaluation method was used to analyze the proposed architecture, perceive its excellence in accuracy, feature optimization, and reliability, specifically for the pothole detection problem.

3. Results and Discussions

To evaluate the performance of the proposed integrated network, classification outcomes are analyzed on the testing dataset, which comprises two classes. Images without potholes are labeled Class 0, while those containing potholes are labeled Class 1 in both the training and testing datasets. The performance of the proposed network is assessed for each class individually using several evaluation metrics. Detailed results and a comprehensive analysis of the model's performance evaluation process are presented in the subsequent sections.

3.1. MRFO-Efficient-Net

The MRFO with baseline Efficient-Net (B0, B1, B2) is assessed for performance as shown in Figure 8. The accuracy of the models is observed by plotting training and validation curves generated in 100 epochs. The stable training and slightly decreasing validation curve indicate the stability of the architecture during the training process, while suggesting that the model can learn complex patterns without overfitting. The sudden drop in validation accuracy at epoch 63 is likely due to random variation in the validation process rather than a genuine decline in model performance, as evidenced by the immediate recovery in subsequent epochs. The consistent performance across training and validation sets helps instill confidence in the model's ability to generalize new, unseen data.

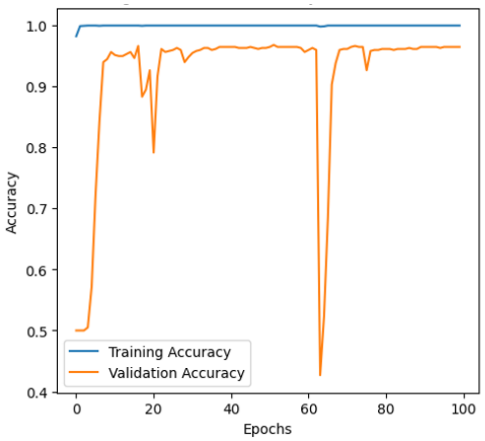
Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421

Additionally, it highlights the importance of carefully monitoring these metrics to ensure the training process remains on track and achieves the desired performance goals.

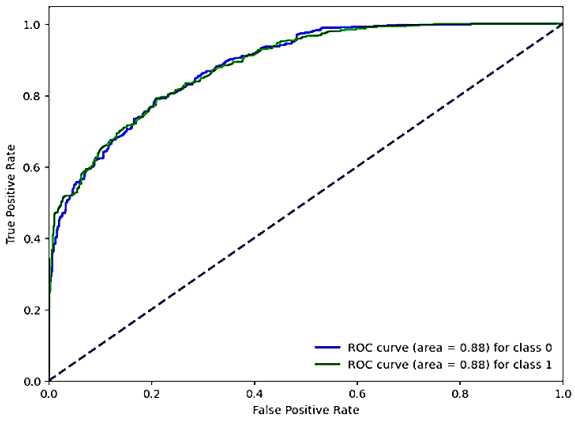
The integrated MRFO with baseline Efficient-Net B0 successfully identified class 0, normal road images, and class 1 potholes. The upward trend observed in the ROC curve for both classes, with the false-positive rate represented on the X-axis and the true-positive rate on the Y-axis, signifies a high degree of accuracy. Furthermore, the integrated MRFO model, in conjunction with the baseline Efficient-Net B0, exhibits a commendable capacity to differentiate between classes effectively.

The MRFO-Efficient-Net B1, due to its slightly complex architecture, achieved better accuracy in 100 epochs. Initially, during the first 30–35 epochs, the model reflected overfitting. However, with increased training, a high training accuracy of above 97% and a testing accuracy of 98% were achieved for normal surfaces and potholes, respectively, as obvious from the training and validation curves and AUC.

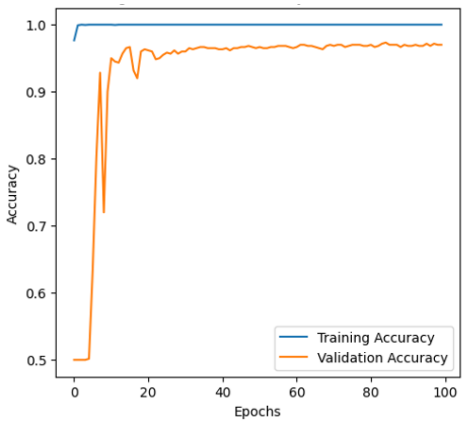
The proposed MRFO-based optimization significantly improved the performance of the Efficient-Net family models for pothole detection. The most complex architecture, Efficient-Net B2-MRFO, achieved the highest classification rates for both normal surfaces and pothole detection. These results indicate that the integration of MRFO aided progressively better optimization of hyperparameters, which led to improved generalization and reduced misclassification rates across all Efficient-Net variants.



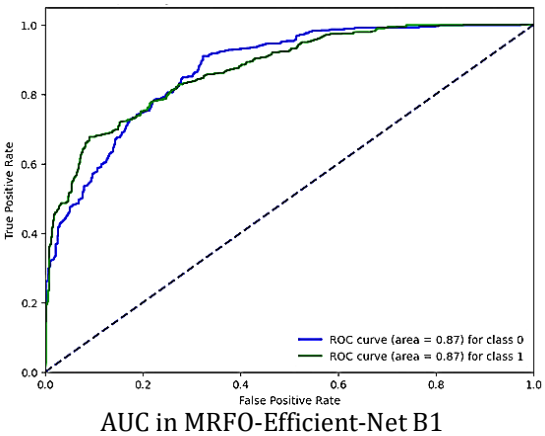
Training and validation accuracy of MRFO-Efficient-Net B0



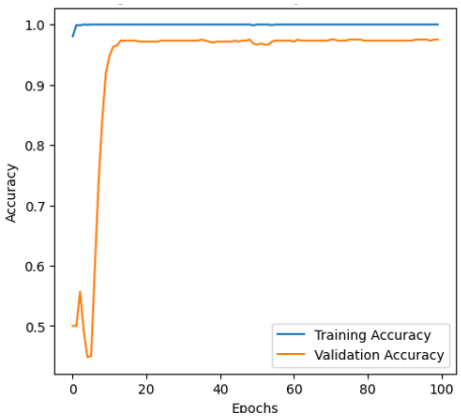
AUC in MRFO-Efficient-Net B0



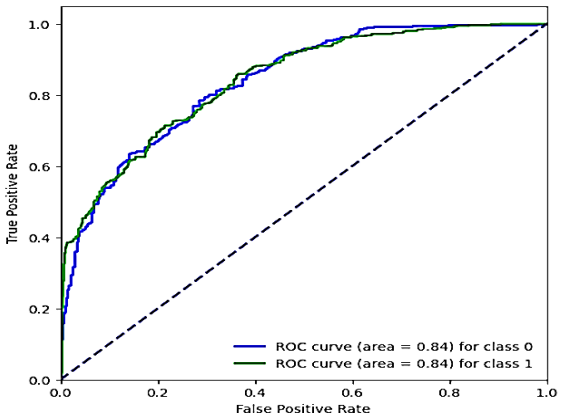
Training and validation accuracy of MRFO-Efficient-Net B1



AUC in MRFO-Efficient-Net B1



Training and validation accuracy of MRFO-Efficient-Net B2



AUC in MRFO-Efficient-Net B2

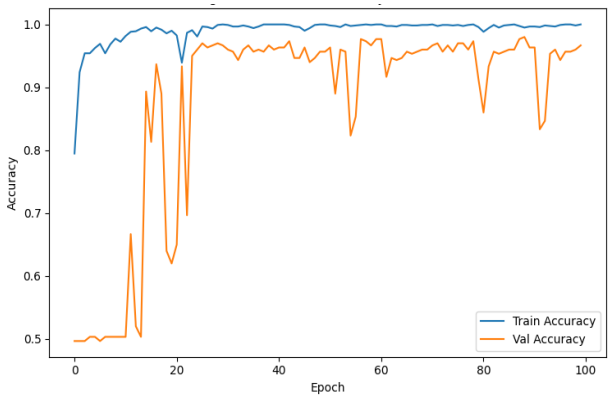
Fig. 8 *Training and validation curves and AUC for MRFO-Efficient-Net (B0-B2)*

3.2. AO-Efficient-Net

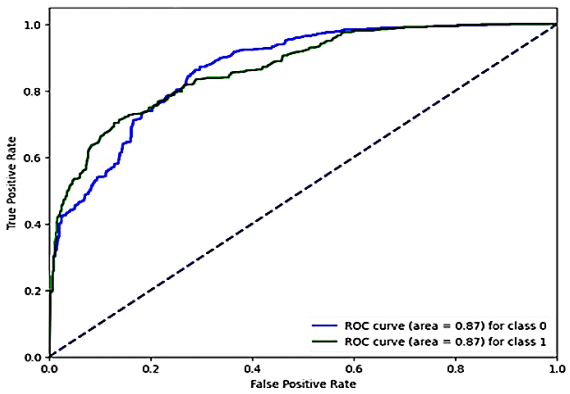
The AO-Efficient-Net B0 model was evaluated using learning curves and AUC shown in Figure 9. It achieved 86% accuracy for normal surfaces and 64% for pothole detection, indicating moderate classification performance. The lower accuracy, particularly for potholes, suggests that the model did not train well on images with complex variations in pothole features and requires further refinement. While AO-Efficient-Net B0 performs adequately in some scenarios, its optimization approach does not match the robustness needed for more diverse datasets. Advanced optimization strategies could potentially enhance its ability to generalize and improve detection accuracy.

The AO-Efficient-Net B1 model achieved 88% accuracy for normal detection and 76% for pothole detection. This represents a clear improvement over AO-Efficient-Net B0, especially for potholes (an increase from 64% to 76%). The deeper architecture of B1 likely allows it to capture more complex features of road surfaces. However, the training and validation curves display a continuous downward trend, indicating overfitting when the model learns training data well but struggles to maintain performance on unseen data. Such overfitting may stem from insufficient regularization, limited dataset size, or an architecture that is overly complex relative to the available data. While the accuracy results are promising, the curve patterns point to a decline in generalization capability.

The AO-Efficient-Net B2 model maintained 88% accuracy for normal detection and 76% for pothole detection, results almost identical to B1. This indicates that increasing the architecture complexity from B1 to B2 did not yield meaningful performance gains for this task. While normal surface detection is strong, pothole detection performance continues to lag. It reflects the inherent difficulty in pothole features. AUC values are reasonably high, indicating good but not perfect class separability. The training and validation curves reveal instability, with oscillations persisting over 100 epochs rather than converging smoothly, exhibiting challenges in achieving consistent learning.

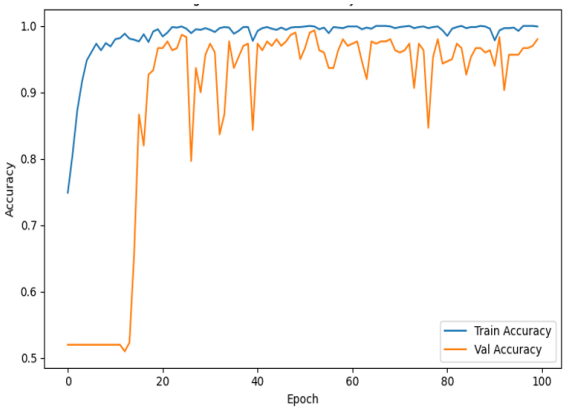


Training and validation accuracy of AO-Efficient-Net B0

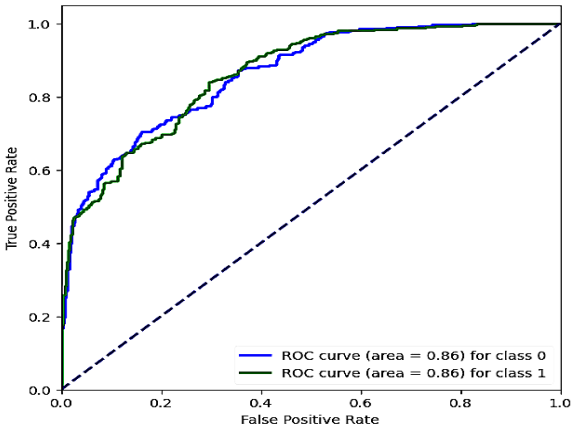


AUC in AO-Efficient-Net B0

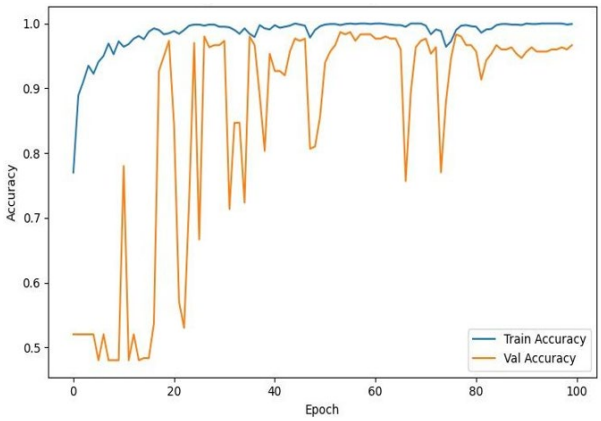
Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421



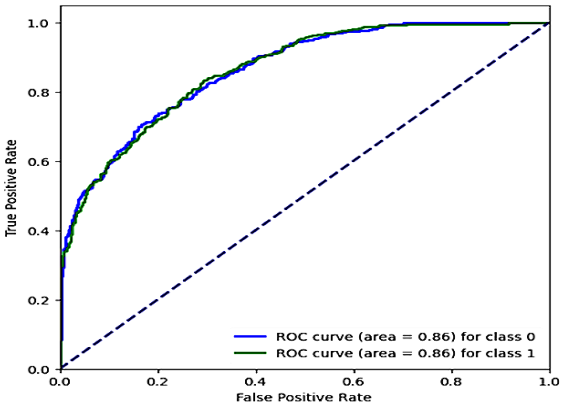
Training and validation accuracy of AO-Efficient-Net B1



AUC in AO-Efficient-Net B1



Training and validation accuracy of AO-Efficient-Net B2



AUC in AO-Efficient-Net B2

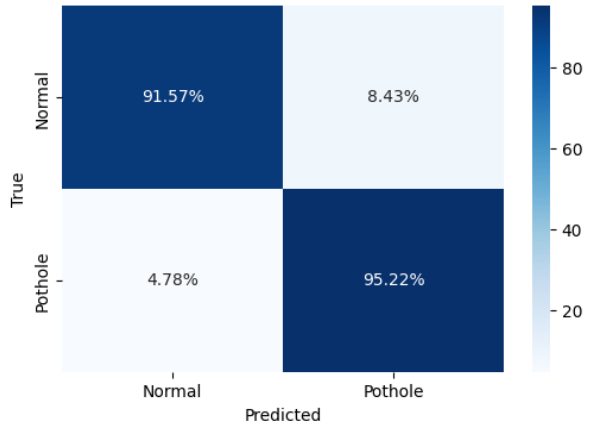
Fig. 9 *Training and validation curves and AUC for AO-Efficient-Net (B0-B2)*

3.3. Confusion Matrix Analysis of MRFO and AO

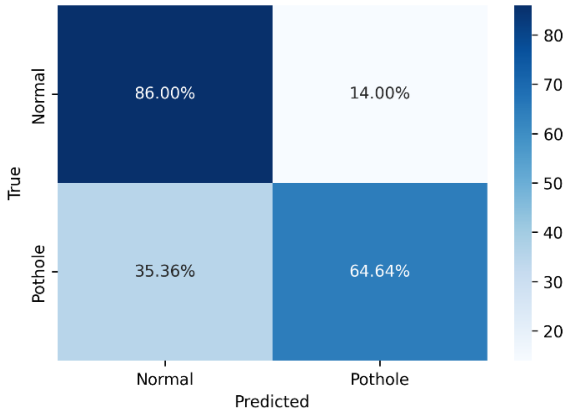
The confusion matrices for all Efficient-Net variants integrated with MRFO and AO reveal clear differences in classification performance for normal surfaces and pothole detection. For the MRFO-based models, performance improved consistently from B0 to B2. MRFO–Efficient-Net B0 achieved a classification accuracy of 91.57% for normal surfaces and 95.22% for pothole detection. In contrast, the same architecture, Efficient-Net B0, with AO could achieve 86% accuracy for normal roads while only 64.64% for potholes. Similarly, there is a significant difference between MRFO-Efficient-Net B1 and AO-Efficient-Net B1, with 93.43% and 86.26% respectively, for normal surface images. For pothole detection, MRFO-Efficient-Net B1 achieved 97.22% accuracy and outperformed AO-Efficient-Net B1 with 75.11% accuracy. Meanwhile, MRFO-Efficient-Net B2 proved itself the best architecture after reaching 97.52% and 98.32% for normal road and pothole detection, respectively.

In contrast, the AO-based model showed notably lower accuracy, particularly for pothole detection. The highest rates of the most complex architecture, MRFO-Efficient-Net B2, demonstrate the optimizer's ability to fine-tune hyperparameters effectively and enhance generalization. This progression indicates that MRFO optimization not only reduces misclassification rates but also scales well with increasing model complexity.

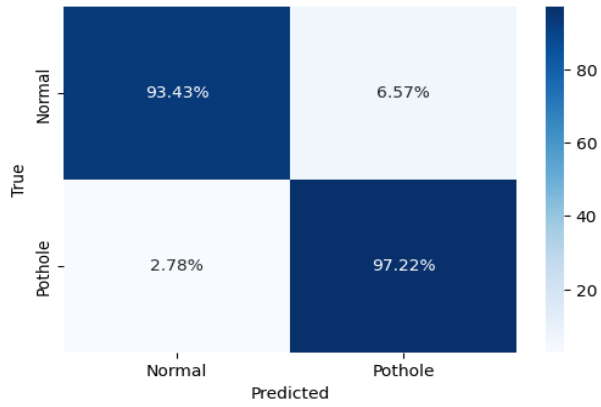
When directly comparing the best-performing variants of each optimizer, MRFO–Efficient-Net B2 achieved 97.52% accuracy for normal surfaces and 98.32% for pothole detection, while AO–Efficient-Net B2 lagged at 88.89% and 76.47%, respectively. This substantial gap is due to the MRFO's superior capability in achieving better class separation and reducing misclassifications, especially for the more challenging pothole category.



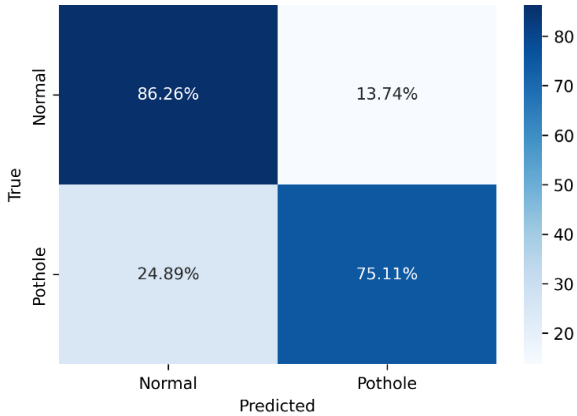
Confusion Matrix % of MRFO-Efficient-Net B0



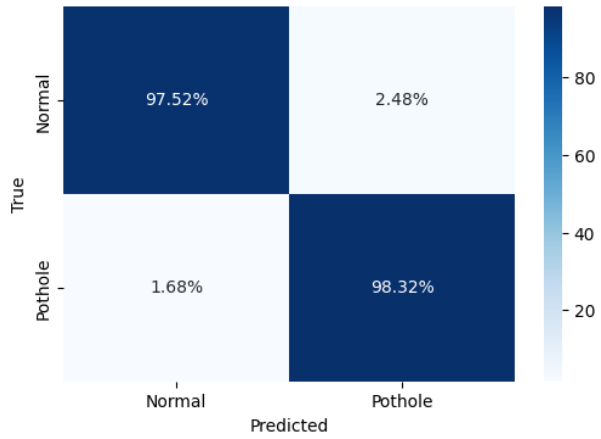
Confusion Matrix % of AO-Efficient-Net B0



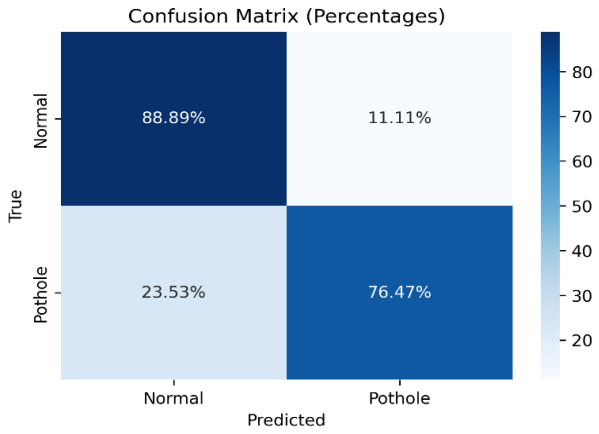
Confusion Matrix % of MRFO-Efficient-Net B1



Confusion Matrix % of AO-Efficient-Net B1



Confusion Matrix % of MRFO-Efficient-Net B2



Confusion Matrix % of AO-Efficient-Net B2

Fig. 10 *Confusion Matrix of MRFO- Efficient-Net (B0-B2) and AO-Efficient-Net (B0-B2)*

3.4. Comparison of MRFO and AO Using Evaluation Metrics

To complement the confusion matrix analysis, precision, recall, and F1-score were calculated for MRFO-Efficient-Net and AO-Efficient-Net with B0, B1, and B2 architectures. These metrics provide a more detailed evaluation of the ability of models to correctly classify instances, focusing on both the accuracy of positive predictions (precision) and the completeness of correct detections (recall).

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 402-421

The outcomes obtained by calculating the accuracy, Precision, Recall, and F1-score of all architectures are reported in Table 2. In six architectures evaluated with four different metrics, MRFO-Efficient-Net outperformed AO-Efficient-Net in accuracy, precision, recall, and F1-score, indicating superior capability in balancing false positives and false negatives. MRFO-Efficient-Net B2 achieved the highest F1-score overall with strong and consistent classification performance.

The improvement resulting from MRFO integration is evident in all metrics. For example, MRFO improves accuracy over AO by 30.58 % for B0, 22.11 % for B1, and 21.85 % for B2. Precision and recall also show marked increases, suggesting that MRFO not only reduces false positives but also captures a higher proportion of true positives. The F1-score improvements confirm that the gains are balanced between precision and recall. These results indicate that MRFO provides a more effective parameter search strategy for Efficient-Net training than AO, likely due to a better exploration–exploitation balance. The well-balanced exploration and exploitation help avoid premature convergence and obtain more accurate model generalization. The consistent superiority of MRFO for all Efficient-Net variants suggests its robustness for optimizing deep CNN architectures.

The deeper and wider Efficient-Net B2, with its higher resolution, achieved a satisfactory accuracy of 76% when optimized with AO, and its performance improved to 98% with MRFO. The 22% increase in classification accuracy was achieved by the MRFO algorithm, which conducted a more efficient global search than the traditional gradient-based method AO. Through a dynamic balance, MRFO assisted the network in steering clear of suboptimal local minima, leading to convergence towards globally optimal solutions. The optimization approach directly influences the improved generalization performance noted in image classification tasks. These findings emphasize the superior performance of the MRFO-optimized Efficient-Net architecture and point out the importance of selecting the best Efficient-Net configuration. Moreover, the results highlight the critical role of optimization algorithms in enhancing image processing and pothole detection accuracy. This comparison highlights the critical role of optimization in advancing pothole detection models. By optimizing the training process, significant improvements in model performance and accuracy can be achieved.

Table 2 *Performance comparison of Efficient-Net architectures integrated with AO and MRFO*

Model	Optimizer	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Efficient-Net B0	AO	64.64	63.20	65.10	64.14
Efficient-Net B0	MRFO	95.22	91.87	95.22	93.49
Efficient-Net B1	AO	75.11	74.60	75.50	75.05
Efficient-Net B1	MRFO	97.22	93.51	97.22	95.29
Efficient-Net B2	AO	76.47	76.00	76.90	76.45
Efficient-Net B2	MRFO	98.32	97.54	98.32	97.88

Optimization is essential for enhancing the efficiency and reliability of pothole detection systems, and the proposed generic architecture can also be utilized in other real-world applications. The experimental results clearly demonstrate the superior performance of Efficient-Net architectures when optimized using MRFO compared to AO. Several technical factors contribute to this outcome. Firstly, MRFO is a metaheuristic algorithm inspired by the foraging behavior of manta rays. It excels at exploring the hyperparameter search space more thoroughly than gradient-based optimizers. During the early phases of training, MRFO samples a wide range of hyperparameter combinations to avoid premature convergence to suboptimal regions of the solution space. As training progresses, MRFO gradually transitions from global exploration to local exploitation. It refines the hyperparameter values around high-performing configurations. This adaptive search strategy results in more stable convergence and better generalization performance across validation datasets.

In contrast, the AO, which updates model weights during training, is not inherently designed for global hyperparameter optimization. It relies on manually selected or grid-searched hyperparameter values (learning rate, beta coefficients, dropout rates), which can limit its adaptability, especially in complex tasks such as pothole detection. Because in these scenarios, data conditions are highly variable due to lighting, shadows, and textures. Pothole detection requires precise feature extraction to tackle these challenging visual variances. MRFO's ability to fine-tune critical hyperparameters such as learning rate, convolutional filter sizes, dropout rates, and dense layer units enhances the feature representation capacity in Efficient-Net. Consequently, MRFO-optimized models exhibited smoother convergence curves, faster training, and higher AUC scores regardless of Efficient-Net variants (B0, B1, B2).

Moreover, since Efficient-Net relies on compound scaling of depth, width, and resolution, MRFO's capacity to simultaneously optimize multiple hyperparameters in a high-dimensional search space becomes especially valuable. Unlike AO, which focuses only on optimizing weights through backpropagation, MRFO performs a broader architectural-level search. This gives it a significant advantage in fine-tuning models for real-world

applications like pothole detection, where both accuracy and efficiency are critical. Therefore, the overall impact of MRFO indicated an improved and robust model.

3.5. Comparison of MRFO-Efficient-Net with Architectures from Existing Literature

To further validate the performance of the proposed MRFO-Efficient-Net architecture, we compared its accuracy with existing studies. For a fair comparison, we consider our highest-accuracy model, MRFO-Efficient-Net B2, with the models achieving the highest accuracy reported in other studies. The studies being compared utilized the same dataset to train different architectures. While earlier sections have demonstrated that our approach outperforms both AO-based optimization and the baseline Efficient-Net, this additional comparison validates the robustness of the proposed model. As it consistently delivers superior or competitive results against previously established popular methods, given in Table 3.

Table 2 *Comparison of MRFO–Efficient-Net B2 with existing state-of-the-art models*

Model	Accuracy (%)	Reference
YOLOv4-tiny	78.7	[30]
YOLOv3	77.1	[31]
ResNet-18	75.5	[32]
CNN with Transfer learning	96	[12]
YOLOv4-tiny	90	[33]
Faster RCNN with Inception v2	90	[34]
MRFO-Efficient-Net B2	98.32	Our study

The superior performance of the proposed MRFO-Efficient-Net B2 model is driven by a combination of optimized features and fine-tuned hyperparameters. While earlier models such as YOLO variants and ResNet-18 demonstrate reasonable detection capability, they are either limited by smaller receptive fields, less efficient scaling, or suboptimal training parameter selection. Even the high-performing CNN with transfer learning achieves lower accuracy, likely due to less specialized optimization for the dataset. In contrast, our proposed approach combines architectural efficiency and metaheuristic optimization. It captures more discriminative features to achieve higher accuracy compared to other models. Since the MRFO algorithm outperforms traditional optimizers by simulating adaptive foraging behaviors to dynamically balance global and local search. This ensures a more thorough exploration of the parameter space during early training and precise exploitation during convergence, which reduces the chances of getting trapped in suboptimal minima. While the experimental results are promising, the study has some limitations. The dataset size, although augmented, may not fully capture the variability in real-world road conditions. Moreover, the absence of real-time or embedded deployment testing restricts our understanding of the practical applicability of the proposed architecture. Future work will focus on scaling data collection and deploying the system in real-world driving scenarios.

4. Conclusion

This study introduced an innovative pothole detection framework by integrating Efficient-Net architectures (B0, B1, B2) with the Manta Ray Foraging Optimization (MRFO) algorithm. The model was trained using two classes of images with potholes and non-potholes. The performance of the proposed architecture was rigorously compared against the conventional Efficient-Net architecture optimized using the Adam optimizer (AO). Based on the experimental results, the proposed MRFO-Efficient-Net architecture demonstrated superior performance, achieving accuracy improvements of 30.58%, 22.11%, and 21.85% for Efficient-Net B0, B1, and B2, respectively, compared to their counterparts optimized with AO. Similarly, the architecture proved more accurate and efficient when compared to a baseline CNN model without optimization. When compared with the best-performing models from existing deep learning studies, MRFO- Efficient-Net -B0 achieved superior performance. The proposed Efficient-Net -MRFO framework consistently outperformed both alternatives in terms of accuracy, the confusion matrix, and the ROC, showing the tradeoff between true positive rate and false positive rate. All three Efficient-Net architectures demonstrated superior accuracy and significantly improved performance when integrated with MRFO. The experimental results highlight the critical role of optimization algorithms and network complexity in achieving higher accuracy for pothole detection. Furthermore, the choice of a suitable optimizer significantly influences the overall performance of deep learning models. These findings emphasize the importance of balancing model complexity and optimization strategies to enhance real-world applicability. In future work, this study will be extended to incorporate real-time video data to train the proposed architecture for pothole detection

for its deployment in dynamic, real-world environments. Such advancements have the potential to contribute significantly to the development of real-time intelligent transportation systems and proactive road maintenance solutions with enhanced road safety and infrastructure management.

Acknowledgement

This research was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2024/ICT02/UTHM/02/4) Vot K506 and Universiti Tun Hussein Onn Malaysia through GPPS Vot Q068.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design****: Muhammad Numan Ali Khan, Mohd Norzali Haji Mohd, Tasiransurini binti Ab Rahman;* ***data collection:*** *Muhammad Numan Ali Khan, Mohd Norzali, Tasiransurini binti Ab Rahman ;* ***analysis and interpretation of results:*** *Muhammad Numan Ali Khan, Mohd Norzali Haji Mohd, Nuzhat Khan, Muhammed Paend Bakht****; draft manuscript preparation****: Muhammad Numan Ali Khan, Mohd Norzali Haji Mohd, Nuzhat Khan, Muhammed Paend Bakht. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] A. S. Duggal *et al.*, "Infrastructure, mobility and safety 4.0: Modernization in road transportation," *Technology in Society,* vol. 67, p. 101791, 2021.

[2] R. Bhatt, J. Shah, and S. Patel, "Enhancing Transportation Infrastructure: A Deep Learning Approach for Pothole Detection," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2024: IEEE, pp. 1-6.

[3] Y. Safyari, M. Mahdianpari, and H. Shiri, "A review of vision-based pothole detection methods using computer vision and machine learning," *Sensors,* vol. 24, no. 17, p. 5652, 2024.

[4] M. H. Asad, S. Khaliq, M. H. Yousaf, M. O. Ullah, and A. Ahmad, "Pothole detection using deep learning: A real-time and AI-on-the-edge perspective," *Advances in Civil Engineering,* vol. 2022, no. 1, p. 9221211, 2022.

[5] M. Jakubec, E. Lieskovská, B. Bučko, and K. Zábovská, "Comparison of CNN-based models for pothole detection in real-world adverse conditions: Overview and evaluation," *Applied Sciences,* vol. 13, no. 9, p. 5810, 2023.

[6] T. Wu, H. Zhu, H. Fan, and H. Zhou, "An improved target detection algorithm based on EfficientNet," in *Journal of Physics: Conference Series*, 2021, vol. 1983, no. 1: IOP Publishing, p. 012017.

[7] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*, 2021: PMLR, pp. 10096-10106.

[8] S. R. Dubey, S. Chakraborty, S. K. Roy, S. Mukherjee, S. K. Singh, and B. B. Chaudhuri, "diffGrad: an optimization method for convolutional neural networks," *IEEE transactions on neural networks and learning systems,* vol. 31, no. 11, pp. 4500-4511, 2019.

[9] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," *Engineering Applications of Artificial Intelligence,* vol. 87, p. 103300, 2020.

[10] S. Adamu *et al.*, "Unleashing the power of Manta Rays Foraging Optimizer: A novel approach for hyper-parameter optimization in skin cancer classification," *Biomedical Signal Processing and Control,* vol. 99, p. 106855, 2025.

[11] J. Yu *et al.*, "Road surface defect detection—from image-based to non-image-based: a survey," *IEEE transactions on intelligent transportation Systems,* 2024.

[12] K. A. Vinodhini and K. R. A. Sidhaarth, "Pothole detection in bituminous road using CNN with transfer learning," *Measurement: Sensors,* vol. 31, p. 100940, 2024.

[13] G. Rangel, J. C. Cuevas-Tello, J. Nunez-Varela, C. Puente, and A. G. Silva-Trujillo, "A survey on convolutional neural networks and their performance limitations in image recognition tasks," *Journal of sensors,* vol. 2024, no. 1, p. 2797320, 2024.

[14] A. K. Pandey, R. Iqbal, T. Maniak, C. Karyotis, S. Akuma, and V. Palade, "Convolution neural networks for pothole detection of critical road infrastructure," *Computers and Electrical Engineering,* vol. 99, p. 107725, 2022.

[15] T. Huang, X. Huang, and H. Yin, "Deep learning methods for improving the accuracy and efficiency of pathological image analysis," *Science Progress,* vol. 108, no. 1, p. 00368504241306830, 2025.

[16] M. M. Islam, M. A. Talukder, M. A. Uddin, A. Akhter, and M. Khalid, "Brainnet: precision brain tumor classification with optimized efficientnet architecture," *International Journal of Intelligent Systems,* vol. 2024, no. 1, p. 3583612, 2024.

[17] M. N. A. Khan, M. N. H. Mohd, F. S. Khan, T. B. Ab Rahman, U. U. Sheikh, and M. P. Bakht, "FedUNA: A Federated Learning Approach for Robust and Privacy-Preserving Pothole Classification using EfficientNet," in *2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA)*, 2024: IEEE, pp. 1-6.

[18] A. Aljohani, "Optimized Convolutional Forest by Particle Swarm Optimizer for Pothole Detection," *International Journal of Computational Intelligence Systems,* vol. 17, no. 1, p. 7, 2024.

[19] P. O. Odion and K. B. Ishola, "Multi-Class Road Defects Detection and Classification System Using Transfer Learning-Based Deep Convolutional Neural Networks," *International Journal of Science for Global Sustainability (IJSGS),* vol. 10, no. 2, 2024.

[20] I. D. Pratama, H. a. Mahmudah, and R. W. Sudibyo, "Design and implementation of real-time pothole detection using convolutional neural network for IoT smart environment," in *2021 International Electronics Symposium (IES)*, 2021: IEEE, pp. 675-679.

[21] V.-T. Hoang and K.-H. Jo, "Practical analysis on architecture of EfficientNet," in *2021 14th International Conference on Human System Interaction (HSI)*, 2021: IEEE, pp. 1-4.

[22] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114.

[23] A. Kumar, "Pothole Detection Dataset," in *kaggle,* ed, 2025.

[24] A. Dhiman and R. Klette, "Pothole detection using computer vision and learning," *IEEE Transactions on Intelligent Transportation Systems,* vol. 21, no. 8, pp. 3536-3550, 2019.

[25] H. Yao, W. Yu, and X. Wang, "A feature memory rearrangement network for visual inspection of textured surface defects toward edge intelligent manufacturing," *IEEE Transactions on Automation Science and Engineering,* vol. 20, no. 4, pp. 2616-2635, 2022.

[26] B.-F. Wu, Y.-C. Wu, L.-W. Chiu, and H.-P. Liu, "Soft label with channel encoding for dependent facial image classification," *IEEE Access,* vol. 10, pp. 10661-10672, 2022.

[27] K. Zhang, Y. Liu, X. Wang, F. Mei, H. Kang, and G. Sun, "IBMRFO: Improved binary manta ray foraging optimization with chaotic tent map and adaptive somersault factor for feature selection," *Expert Systems with Applications,* vol. 251, p. 123977, 2024.

[28] K. Ramyashree, B. Sharada, and R. Bhairava, "Classification of Human Peripheral Blood Cells Using EfficientNet B0-B7 Models," in *2024 International Conference on Smart Electronics and Communication Systems (ISENSE)*, 2024: IEEE, pp. 1-6.

[29] C. Lin, P. Yang, Q. Wang, Z. Qiu, W. Lv, and Z. Wang, "Efficient and accurate compound scaling for convolutional neural networks," *Neural Networks,* vol. 167, pp. 787-797, 2023.

[30] V.-T. T. Sung-Sik Park, "Application of Various YOLO Models for Computer Vision-Based Real-Time Pothole Detection," *Applied Sciences,* 2021.

[31] M. Jakubec, "Comparison of CNN-Based Models for Pothole Detection in Real-World Adverse Conditions: Overview and Evaluation," *Applied Sciences,* 2023.

[32] D. B. Naik, "Pothole Detection and Classification in Vehicular Networks using ResNet-18," presented at the 2023 IEEE 7th Conference on Information and Communication Technology (CICT-2023), 2023.

[33] S. K. Muhammad Haroon Asad, "Pothole Detection Using Deep Learning: A Real-Time and AI-on-the-Edge Perspective," *Advances in Civil Engineering Volume 2022, Article ID 9221211, 13 pages,* 2022, doi: https://doi.org/10.1155/2022/9221211.

[34] S. M. Dympna O'Sullivan, "Pothole Detection under Diverse Conditions using Object Detection Models," in *International Conference on Image Processing and Vision Engineering*, 2021: SCITEPRESS, doi: 10.5220/0010463701280136.