INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 08 | AUG - 2025 SJIF RATING: 8.586 ISSN: 2582-3930

# A Review and Analysis of Evolutionary Mathematical Algorithms for Pavement Condition Prediction

Himanshu Kumar Pandey[1], Prof. Vinay W. Deulkar [2]
Department of Civil Engineering,[1,2]
JIT, Borawan, India.[1,2]

***Abstract*—** Pavement infrastructure plays a critical role in the economic development and transportation systems of any country. Maintaining pavements in good condition is essential for ensuring road safety, reducing vehicle operating costs, and prolonging the lifespan of roads. Traditional methods for predicting pavement deterioration are often limited in their ability to handle complex, non-linear relationships and uncertainty in environmental and traffic conditions. Therefore, there is a growing need for advanced computational techniques that can model such intricacies more accurately. Evolutionary mathematical algorithms, inspired by biological evolution and natural selection, offer a promising solution. These algorithms, including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Differential Evolution (DE), and Evolutionary Strategies (ES), are capable of exploring large solution spaces efficiently. They are well-suited for optimizing complex systems where traditional methods fall short. Their ability to evolve solutions over iterations allows them to adapt to the changing patterns and uncertainties inherent in pavement condition data. This paper presents a comprehensive review on evolutionary mathematical algorithms for pavement condition prediction along with their salient features.

***Keywords*— *Pavement Condition Prediction, Evolutionary Models, Swarm Intelligene, Classification Accuracy.***

## I. INTRODUCTION

Conventional statistical models such as regression and deterministic equations rely on historical data and fixed relationships among variables, which may not fully capture the dynamic and stochastic nature of pavement degradation. These models often struggle with issues like data sparsity, variability in material behavior, and unanticipated climate influences. Furthermore, their predictive accuracy diminishes when dealing with non-linear and multi-objective optimization problems, which are common in infrastructure systems. This limitation highlights the demand for more flexible and adaptive approaches [1].

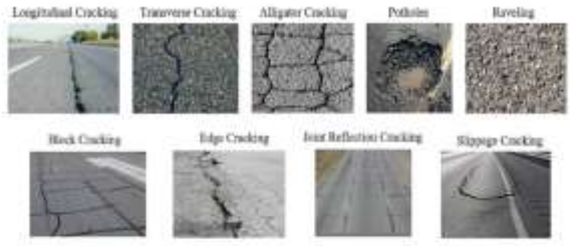


**Fig.1 Common Pavement Failures.**

Figure 1 depicts the common pavement failure categories. With the advancement of smart transportation and infrastructure monitoring systems, the integration of evolutionary algorithms has become increasingly feasible and impactful. They can be embedded within Internet of Things (IoT)-enabled platforms for real-time pavement monitoring and predictive maintenance scheduling. Such integration not only enhances prediction accuracy but also enables proactive maintenance strategies, leading to cost savings and improved public safety [2].

## II. IMBALANCED DATASETS

Pavement condition prediction is a crucial task in infrastructure management, helping agencies allocate maintenance resources effectively and ensure road safety. However, the accuracy of predictive models heavily depends on the quality and distribution of the input data. One major challenge in this domain is the problem of imbalanced datasets, where the number of samples across different pavement condition classes [3]:

1. Good
2. Fair
3. Poor

The samples of survey are generally not uniformly distributed. This imbalance poses significant obstacles to the development of robust and generalizable predictive models.

In pavement condition datasets, imbalance typically arises because the majority of roads are often in acceptable or 'Good' condition due to regular maintenance and upgrades, while only a small fraction fall into the 'Poor' category. As a result, data collected from real-world surveys or monitoring systems naturally exhibit class imbalance. Other contributing factors include biased data sampling, insufficient historical records of degraded pavements, and lack of data from remote or lesstraveled roads. This skew in class representation leads to an underrepresentation of critical degradation patterns [4].

This imbalance can skew the model's performance, making it less effective at identifying pavement condition. Techniques such as oversampling, undersampling, and synthetic data generation are often employed to address this issue. The Skewness **(Pearson's Moment Coefficient of Skewness)** is defined as [5]:

$$S = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{sd}\right)^3 \quad (1)$$

Here,
$n$ is number of observations.
$x_i$ is each individual data point.
$\overline{x}$ is the mean.
$sd$ denotes standard deviation

One of the most effective techniques is the random oversampling method mathematically described as:

In this case, we need to add copies or synthetic data to minority classes. For minority class $C_j$ , generate new samples until [6]:

$$N'_j = N_{major} = \max(N_i) \quad (2)$$

Thus the new distribution becomes:

$$P' = \left[\frac{N'_1}{N'}; \frac{N'_2}{N'}, \frac{N'_k}{N'}\right] \approx Uniform \quad (3)$$

The above can be done through cost sensitive learning based on the following formulation:
Assign different misclassification costs to classes to adjust bias. Let the cost matrix be $C = [c_{ij}]$, then, $c_{ij}$ denotes the cost of predicting class $j$ when the true class is $i$.

$$c_{ii} = 0; i = j \quad (4)$$

$$c_{ij} = \frac{1}{N_i}; i \neq j \quad (5)$$

In such a case, the loss function becomes [7],

$$L = \sum_{i=1}^{N} C(y_i, \tilde{y}_i) * L(f(x_i) y_i) \quad (6)$$

This penalizes errors more in underrepresented classes, mathematically correcting skewness during model training. It can be implemented through an adaptive approach to update system weights. A swarm intelligence based approach can be effective in this regard.

## III. COMMON EVOLUTIONARY MODELS

**Decision trees:** Decision trees are commonly used for fraud detection since they are straightforward and easy to understand. They operate by partitioning the dataset into subsets according to the input feature values, resulting in a hierarchical structure of decision nodes. Every node in the representation reflects a specific feature, each branch represents a decision rule, and each leaf represents an outcome. Decision trees possess the capability to process both numerical and categorical input, rendering them adaptable and comprehensible [8].

**Random Forests:** Random forests use the idea of decision trees by employing a collection of numerous trees to enhance accuracy and resilience. The construction of each tree in a random forest involves using a random subset of the data, which helps to reduce overfitting and improve prediction performance. Random forests excel at detecting intricate fraud patterns in extensive datasets, providing exceptional accuracy and robustness against interference.

**Logistic regression:** Logistic regression is a statistical model that is specifically designed for binary classification tasks, allowing it to effectively differentiate between fraudulent and non-fraudulent transactions. The logistic function is used to evaluate the probability of a given input belonging to a specific class. Logistic regression is renowned for its simplicity, efficiency, and interpretability. It is particularly useful in cases where the relationships between features may be approximated as linear.

**Support Vector Machines (SVM):** Support vector machines (SVM) are robust classifiers that identify the most effective hyperplane for distinguishing between various classes in a space with many dimensions. Support Vector Machines (SVMs) are highly efficient in dealing with data that has a large number of dimensions. They are particularly valuable when the classes cannot be separated by a straight line, as they can employ kernel functions to transform inputs into spaces with even more dimensions. The capacity of SVMs to detect fraud makes them a highly advantageous option.

**Artificial neural networks:** Neural networks, particularly deep learning models, have become popular due to their capacity to acquire intricate patterns from extensive datasets. Neural networks are capable of representing complex connections between characteristics in fraud detection, enabling them to detect tiny deviations that are indicative of fraudulent activity. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are utilized depending on the characteristics of the input and the specific demands of the task.

**K-Nearest Neighbors (KNN):** The K-nearest neighbors (KNN) technique is an instance-based learning method utilized for categorization. It identifies the 'k' most comparable transactions (neighbors) to a particular transaction. The class that has the highest number of occurrences among the neighbors is allocated to the new transaction. KNN is characterized by its simplicity and intuitiveness, yet it may incur high computing costs when dealing with extensive datasets. However, it is still efficient for smaller datasets and can yield rapid, easily understandable outcomes.

***SWARM INTELLIGENCE:***

Swarm Intelligence (SI) is a subfield of artificial intelligence that focuses on the collective behavior of decentralized and self-organized systems, typically inspired by natural phenomena such as bird flocking, fish schooling, or insect swarming. These systems consist of simple agents interacting locally with one another and with their environment, resulting in the emergence of intelligent global behavior. Swarm intelligence has proven to be a powerful paradigm for solving complex optimization problems where traditional methods struggle.

The core principle of swarm intelligence lies in the idea that a group of simple entities can collectively solve difficult tasks without a central coordinator. Each individual agent follows simple rules based on local information, and the group behavior emerges from their interactions. These systems are adaptive, robust, and scalable. Swarm intelligence models often exhibit characteristics such as distributed control, self-organization, and stigmergy (indirect communication through environmental modifications), which make them suitable for dynamic and uncertain environments [9].

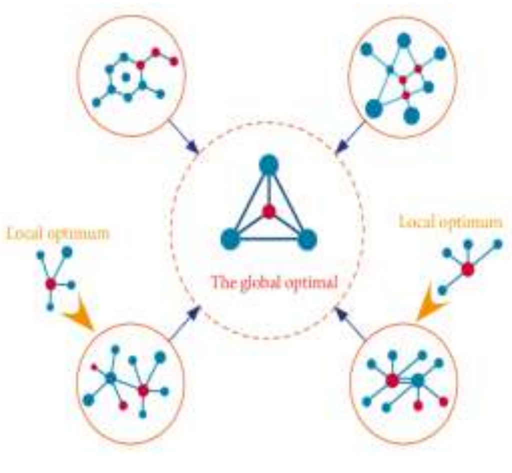


**Fig.2 Visualization of Swarm Intelligence**

Figure 2 depicts the concept of swarm intelligence. Swarm intelligence, and particularly swarm intelligence exemplifies how nature-inspired algorithms can solve complex optimization problems effectively. By leveraging collective behavior and distributed decision-making, the approach efficient framework adaptable to numerous applications. The major swarm intelligence methods happen to be [10]:

1. Particle Swarm Optimization (PSO)
2. Bat Optimization (BO)
3. Ant Colony Optimization (ACO)
4. Teacher Learner Based Optimization (TLBO)

Each of the approaches uses a slightly different training rule. The PSO is governed by:

$$\mathbf{vel_k = w \times vel_{k-1} + L_1\alpha_1[p_i\text{-}x_{i(k-1)}] + L_2\alpha_2[p_g\text{-}x_{i(k-1)}]} \quad \mathbf{(7)}$$

$$\boldsymbol{x_{i(k)} = x_{i(k-1)} + vel_k} \quad \mathbf{(8)}$$

Here,
vel is the particle velocity
k is the iteration
$L_1$ and $L_2$ are learning factor values
$x_i$ is the particle position
$\alpha_1$ and $\alpha_2$ are random number values
w represents the weights
$p_i$ represents particle's individual best position
$p_g$ represents group's best position

## IV. EXISTING WORK

This section presents a review on the baseline approaches in the domain.

**Elshaboury et al. [11]** proposed that Pavement condition prediction helps road agencies to schedule maintenance, rehabilitation, and reconstruction, and to allocate limited funds and resources to such activities.. Additionally, local road pavement condition data may suffer from dataset imbalance, often leading to unreliable condition predictions. Hence, this paper introduces a methodology to predict local pavement condition using various single estimator and ensemble machine learning (ML) models along with the adaptive synthetic sampling method. The study develops nine (9) Bayesian-optimised ML models: category boosting (CatBoost), adaptive boosting, decision tree, extra trees, gradient boosting, light gradient-boosting machine, k-nearest neighbour, random forest, and artificial neural network.

**Ahmed et al. [12]** explored the application of machine learning techniques, including support vector machines and random forests, for predicting the Pavement Condition Index (PCI). The authors compare these methods with traditional regression models, concluding that machine learning offers superior accuracy and reliability in prediction. The study also emphasizes the role of quality input data in improving model performance.

**Zhou et al. [13]** evaluated the effectiveness of multiple machine learning models, including neural networks and ensemble methods, in predicting pavement performance metrics such as roughness and cracking. The study finds that ensemble methods like gradient boosting outperform other models in terms of accuracy and computational efficiency. It highlights the importance of feature engineering for better predictions.

**Gao et al. [14]** focussed on the use of deep learning algorithms, particularly convolutional neural networks (CNNs), for classifying and predicting pavement distresses. By leveraging large datasets, the study demonstrates the effectiveness of deep learning in capturing complex patterns and achieving high accuracy. The research also discusses the challenges of data preprocessing and model interpretability.

**Kim et al. [15]** compared traditional statistical models such as regression with machine learning approaches like random forests and support vector machines for forecasting pavement conditions. The findings indicate that machine learning models provide more robust predictions, especially in scenarios involving non-linear data relationships. The study also highlights the importance of integrating environmental and traffic data for better results.

**Huang et al. [16]** introduced a hybrid machine learning framework combining decision trees and neural networks to predict pavement conditions. The authors show that the hybrid model achieves higher accuracy and reliability than standalone models. The study emphasizes the need for scalable models to handle large infrastructure networks and discusses potential applications in smart city initiatives. The classification accuracy was considered the performance metric.

**Mansorri et al. [17]** evaluated prediction models for the Pavement Condition Index (PCI) using multiple linear regression, artificial neural networks, and fuzzy logic inference models for flexible pavement sections. The authors collected field data spanning from 2018 to 2021, considering eight pavement distress factors as inputs for predicting PCI values. The research demonstrates that machine learning models, particularly artificial neural networks, outperform traditional regression methods in predicting PCI, highlighting the potential of machine learning techniques in pavement condition assessment.

**Kim et al. [18]** presented a machine learning model for predicting surface distresses in concrete pavements, focusing on South Korean examples. By analyzing time series data, the study aims to surpass traditional regression methods in forecasting pavement conditions. The findings suggest that machine learning models can effectively predict various types of pavement distresses, contributing to more sustainable pavement management practices.

**Li et al. [19]** utilized machine learning algorithms to analyze and predict pavement performance, leveraging large datasets like the Long-Term Pavement Performance (LTPP) database. This study focuses on predicting spalling in concrete pavements using statistical and machine learning models. The findings indicate that machine learning models, particularly ensemble methods, provide superior predictive capabilities for pavement distress, emphasizing the value of large datasets in model development.

**Wang et al. [20]** proposed the RF and XGBoost models for predicting pavement condition. he objective of this paper is to develop machine learning models—Random Forest (RF) and eXtreme Gradient Boosting (XGBoost)—to predict pavement surface curvature. The authors demonstrate that these models can effectively assess pavement structural condition, providing a valuable tool for infrastructure maintenance planning. The study highlights the potential of ensemble machine learning methods in enhancing the accuracy of pavement condition assessments. The TP, TN, FP and FN rates were evaluated for the performance evaluation.

## V. CHALLENGES AND EVALUATION METRICS.

Pavement conditions often exist on a continuum, but classification tasks require discretizing these continuous degradation levels into fixed labels. This introduces overlapping class boundaries, where data points near class thresholds (e.g., Pavement Condition Index [PCI] of 70–80) may exhibit ambiguous characteristics. Such cases lead to confusion between adjacent classes, reducing model precision and increasing misclassification rates. Moreover, subjective assessments during data labeling (e.g., visual inspection) further contribute to inconsistencies in class definitions [21].

Data quality significantly influences classification performance. Pavement datasets collected via sensors, cameras, or human inspections often suffer from noise, missing values, and inconsistencies. Environmental factors like lighting, shadows, and weather conditions can degrade image-based data, while sensor malfunctions or reporting errors may lead to inaccurate condition records. These imperfections propagate through the model training process, leading to unreliable classifications and reduced generalization capability [22].

Pavement degradation is influenced by spatial factors (soil type, traffic volume, drainage) and temporal factors (seasonal weather patterns, maintenance history) [23]. Models that fail to incorporate this spatio-temporal variability often struggle with generalization across regions or time periods. For example, a classifier trained on urban road data may not perform well in rural or coastal environments. Building generalized models that account for such heterogeneity remains a major challenge [24].

The overall performance metrics are mathematically defined as [25]:

**Accuracy:** It is mathematically defined as:

$$Ac = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

**Recall:** It is mathematically defined as:

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

**Precision:** It is mathematically defined as:

$$Precisiosn = \frac{TP}{TP+FP} \tag{11}$$

**F-Measure:** It is mathematically defined as:

$$F - Measure = \frac{2.Precision.Recall}{Precision+Recall} \quad (12)$$

Here.
TP represents true positive
TN represents true negative
FP represents false positive
FN represents false negative

**CONCLUSION: In conclusion, the evolving demands of pavement infrastructure management necessitate the adoption of intelligent, adaptive, and robust prediction models. Evolutionary mathematical algorithms offer a powerful toolset for addressing the complexities and uncertainties associated with pavement condition prediction. Their integration with modern technological ecosystems promises a future of smarter, more efficient, and more resilient transportation infrastructure. However, pavement condition classification is fraught with challenges ranging from imbalanced datasets and overlapping class boundaries to real-time constraints and lack of interpretability. Overcoming these obstacles requires a combination of robust data preprocessing, algorithmic innovation, and domain-aware modeling strategies. Future research must focus on developing hybrid models that incorporate spatial-temporal intelligence, handle uncertainty effectively, and align closely with real-world maintenance needs. Only then can classification systems truly enhance the sustainability and safety of roadway infrastructure.**

## References

1. Chen, C., & Wang, K. C. P. (2017). Pavement distress detection using deep learning: A survey. *Journal of Transportation Engineering, Part B: Pavements*, 143(6), 04017027.
2. El-Basyouny, M., & Sayed, T. (2016). Pavement condition prediction using machine learning algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, 2547(1), 97–106.
3. Gharaibeh, N. G., & Alhasan, A. (2018). Predicting pavement condition index using machine learning methods and ensemble techniques. *International Journal of Pavement Engineering*, 19(10), 875–883.
4. Huang, B., Xu, Q., & Wang, K. C. P. (2018). Automated pavement distress detection using deep convolutional neural networks. *International Journal of Pavement Engineering*, 19(5), 457–465.
5. Jiao, J., & Zhang, X. (2016). Pavement distress detection and classification using deep learning. *International Journal of Pavement Engineering*, 17(6), 505–514.
6. Jiang, X., & Adeli, H. (2017). Pavement distress prediction using an ensemble machine learning approach. *Journal of Transportation Engineering, Part B: Pavements*, 143(6), 04017026.
7. Khasnobish, A., & Mukherjee, A. (2018). Pavement condition assessment using machine learning algorithms and fuzzy logic. *Journal of Transportation Engineering, Part B: Pavements*, 144(2), 04018020.
8. Li, Q., & Cui, L. (2017). Pavement performance prediction based on support vector machine and particle swarm optimization. *Journal of Computing in Civil Engineering*, 31(6), 04017056.
9. Shahin, M. Y., & Kohn, S. D. (2016). Pavement condition prediction using artificial neural networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2547(1), 71–78.
10. Wang, T., & Li, Z. (2017). Pavement condition prediction using long short-term memory neural networks. *Journal of Transportation Engineering, Part B: Pavements*, 144(3), 04018026.
11. Wang, Y., Wei, J., Zhang, Z., Hu, Y., & Xu, Y. (2022). Enhancing local road pavement condition prediction using Bayesian-optimized ensemble machine learning and adaptive synthetic sampling technique. Journal of Infrastructure Preservation and Resilience, 3(1), 1-14.
12. Ahmed, S., Hasan, S., & Rahman, T. (2019).Predicting pavement condition index using machine learning techniques. International Journal of Pavement Engineering, 20(4), 567-578.
13. Zhou, Z., Chen, C., & Wang, T. (2020). Machine learning models for pavement performance prediction: A comparison of neural networks, decision trees, and ensemble methods. Journal of Transportation Engineering, Part B: Pavements, 146(1), 04020002.
14. Gao, L., Wang, Y., & Zhang, H. (2021). Pavement distress classification and prediction using deep learning techniques. Computers in Civil Engineering, 38(3), 123-135.
15. Kim, Y., & Kang, D. (2020). Data-driven approaches for pavement condition forecasting using statistical and machine learning models. Transportation Research Record, 2674(5), 11-21.
16. Huang, Y., Xiao, Y., & Sun, W. (2022). A hybrid machine learning framework for pavement condition prediction. Automation in Construction, 132, 104020.
17. Al-Mansoori, S., & Hassan, R. (2022). Predicting pavement condition index based on the utilization of machine learning techniques: A case study. International Journal of Pavement Research and Technology. Advance online publication.
18. Kim, Y., & Kang, D. (2023). Predicting concrete pavement condition for sustainable management. Sustainability, 16(2), 573.
19. Li, H., & Zhang, Y. (2024). Statistical and machine learning models for predicting spalling in concrete pavements. Scientific Reports, 14, 69999.
20. Wang, Y., & Xu, Y. (2022). Machine learning

models for pavement structural condition assessment. Open Journal of Civil Engineering, 12(3), 345-359.

21. Cai, L., Wu, F., & Lei, D. (2021). Pavement Condition Index Prediction Using Fractional Order GM(1,1) Model. IEEJ Transactions on Electrical and Electronic Engineering, 16(8), 1099–1103.

22. Xiao, M., Luo, R., Chen, Y., & Ge, X. (2023). Prediction Model of Asphalt Pavement Functional and Structural Performance Using PSO–BPNN Algorithm. Construction and Building Materials, 407, Article 133534.

23. Basnet, K. S., & Shrestha, R. N. (2023). Pavement Performance Model for Road Maintenance and Repair Planning: A Review of Predictive Techniques. Digital Transportation and Safety, 2(4), 253–267.

24. Bashar, M. Z., & Torres-Machi, C. (2021). Performance of Machine Learning Algorithms in Predicting the Pavement International Roughness Index. Transportation Research Record, 2675(5).

25. Damirchilo, F., Hosseini, A., Mellat Parast, M., & Fini, E. H. (2021). Machine Learning Approach to Predict International Roughness Index Using Long-Term Pavement Performance Data. Journal of Transportation Engineering, Part B: Pavements, 147(4), Article 04021058