

Article

Data-Driven Prediction of Track Quality Index (TQI): A Comparative Study of Statistical and Ensemble Learning Models—A Case Study of the Kolashin–Podgorica Railway

Ivona Nedevska Trajkova ^{1,*} , Zlatko Zafirovski ¹ , Jelena Dimitrijevic ², Riste Ristov ¹  and Vasko Gacevski ¹

¹ Faculty of Civil Engineering, Ss. Cyril and Methodius University in Skopje, 1000 Skopje, North Macedonia; zafirovski@gf.ukim.edu.mk (Z.Z.); ristov@gf.ukim.edu.mk (R.R.); gacevski@gf.ukim.edu.mk (V.G.)

² Faculty of Civil Engineering and Architecture, University of Niš, 18000 Nis, Serbia; jelena.dimitrijevic@gaf.ni.ac.rs

* Correspondence: nedevska@gf.ukim.edu.mk; Tel.: +389-70742824

Abstract

Maintaining the geometric quality of railway tracks is essential for safe and efficient operations. This study examines the prediction of the Track Quality Index (TQI) using Multiple Linear Regression (MLR) and Random Forest (RF) models, based on inspection data from the mountainous Kolashin–Podgorica railway section. Data collected from 2017 to 2022 were used for model development, while 2024 data served as an independent validation set. The primary contribution is a high-resolution analysis of 20-m homogeneous track units across multiple years, combined with independent validation. The dataset includes precise measurements segmented into 20-m units, covering infrastructure, geometric, operational, and maintenance variables. Model variables were selected for their relevance to track quality and data completeness. Both models were trained on scaled input data and evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). The RF model outperformed MLR, achieving a higher R^2 (0.69 versus 0.57) and an MAE approximately 15% lower. Furthermore, the RF model identified severe localized degradation trends and demonstrated robust screening capability, achieving 86.4% sensitivity in detecting high-risk track segments requiring urgent intervention. These findings highlight the effectiveness of ensemble machine learning methods in reducing prediction errors and enabling proactive, data-driven track maintenance in complex railway networks. This methodology allows railway engineers to identify segments most susceptible to rapid deterioration, supporting more precise scheduling of tamping, renewals, or other targeted interventions. Model outputs can assist managers in prioritizing maintenance activities, optimizing resource allocation, and minimizing unexpected failures, thereby enhancing safety and cost efficiency in daily operations. However, the study is limited by its focus on a single railway corridor and a relatively short observation period with few maintenance events, which may restrict the generalizability of the findings to other lines or maintenance regimes.



Academic Editor: Yang Song

Received: 8 June 2026

Revised: 24 July 2026

Accepted: 25 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Track Quality Index (TQI); railway track geometry; predictive maintenance; Random Forest; multiple linear regression; data-driven modeling; Western Balkans

1. Introduction

The geometric condition of railway track infrastructure plays a critical role in determining the safety, ride comfort, and cost-efficiency of railway transport systems [1–3].

Over time, track geometry deteriorates as a result of cumulative dynamic loading and environmental influences, which necessitates systematic maintenance interventions [4,5]. The Track Quality Index (TQI) serves as a widely recognized metric for quantifying track condition by aggregating deviations in key geometric parameters [6–9]. Typically, TQI includes measures of track alignment, longitudinal level, cross-level, gauge, and twist, calculated as either the standard deviation or mean absolute deviation from ideal values over a defined section. Higher TQI values indicate greater deviations and, therefore, reduced track quality. This composite index enables a comprehensive assessment of track condition and supports the prioritization of maintenance activities.

Research on railway maintenance in the Western Balkans remains limited. In North Macedonia, the Study for the Renewal of Corridor X [10] is among the few comprehensive analyses that address both technical and economic aspects of maintenance and rehabilitation along the Tabanovce–Gevgelija line. This study systematically reviewed track geometry degradation and maintenance activities from 1991 to 2000, highlighting significant budgetary and institutional constraints that continue to affect the current railway network. Beyond this example, there is a notable lack of research in North Macedonia that explicitly focuses on track degradation, modeling, and predictive maintenance.

At the regional level, the South-East Europe Transport Observatory (SEETO) produced the Report on Rail Maintenance on the TEN-T Core Network in the Western Balkans [11]. This report analyzed the railway infrastructure of Albania, Bosnia and Herzegovina, Montenegro, North Macedonia, Kosovo, and Serbia, concluding that nearly 30% of the core network was in poor condition. Maintenance budgets were frequently insufficient or allocated without a systematic approach. These findings underscore the urgent need for systematic maintenance strategies and robust degradation models for railways in the Western Balkans.

Internationally, numerous studies have investigated railway track degradation using both statistical and machine learning approaches. Traditional deterministic and probabilistic models have considered variables such as traffic load, radius of curvature, steep slopes, and tamping cycles to estimate track quality deterioration [12]. Recent research has introduced data-driven and artificial intelligence methods, including artificial neural networks for predicting track geometry degradation [13] and hybrid decision-support models that integrate degradation trends with life-cycle cost analysis. These studies demonstrate the potential of predictive modeling to enhance maintenance planning, while also emphasizing the need for case-specific applications in aging railway networks, particularly within the Western Balkans.

Most existing research on railway track degradation has focused on deterministic or probabilistic models that estimate deterioration in quality based on factors such as traffic load, curvature, and tamping cycles [12]. Although these models offer valuable baseline insights, they often neglect the combined effects of infrastructure conditions, including tunnels, bridges, and steep slopes, as well as operational factors such as traffic intensity and speed restrictions. Recent studies employing artificial neural networks [13] or hybrid decision-support systems have achieved greater predictive accuracy. However, these advanced approaches have primarily been implemented in large railway networks in Western Europe or Scandinavia, where data availability and monitoring systems are significantly more advanced.

In contrast, the Western Balkans lack case-specific applications of predictive models, particularly in mountainous sections with complex geometry and frequent maintenance constraints (summarized in Table 1). This study addresses these gaps by comparing statistical and machine learning approaches for the Kolashin–Podgorica line, providing one of the first data-driven predictive analyses of track quality in the region. The methodol-

ogy includes the collection of historical track inspection data from 2017, 2019, and 2022, extraction of key geometric and operational parameters, and application of both Multiple Linear Regression and Random Forest algorithms for predictive modeling. This approach utilizes available infrastructure databases and field measurements to evaluate and compare the predictive performance of each method, with the findings intended to inform practical maintenance planning.

Table 1. Overview of previous studies and reports on railway track maintenance and degradation modeling, with emphasis on Corridor X and the Western Balkans region. (Source: Author’s elaboration [10–13]).

Approach/Study	Methods Applied	Main Limitation	Contribution of This Study
Krakutovski (2009)—Corridor X Study [10]	Technical & economic analysis; descriptive track degradation trends	No predictive modeling; focused only on rehabilitation costs and activities	Provided systematic data on maintenance/rehabilitation in Macedonia
Jovanović et al. (2015) [12]	Deterministic & probabilistic regression	Limited to traffic & curvature; no ML	Established baseline statistical models
Khajehei et al. (2022) [13]	Artificial Neural Networks	Applied in Sweden; requires large datasets	Demonstrated ML potential for TQI prediction
SEETO Report (2016) [11]	Descriptive infrastructure analysis	No predictive modeling	Identified poor maintenance in the Western Balkans
This study	Multiple Linear Regression & Random Forest	Applied to Kolashin–Podgorica (Western Balkans)	First comparative predictive modeling in a regional context

Accurate forecasting of future Track Quality Index (TQI) values enables railway infrastructure managers to implement proactive and optimized maintenance strategies [14,15]. Traditional statistical methods, such as Multiple Linear Regression (MLR), offer interpretable baseline models but often fail to capture complex nonlinear relationships [16–19]. In contrast, ensemble learning techniques, including Random Forest (RF), provide improved accuracy and robustness, especially in high-dimensional contexts [20–22].

This study evaluates the performance of MLR and RF models in predicting TQI for 2024, utilizing historical data from three previous inspection intervals: 2017, 2019, and 2022. The objective is to identify the most suitable method for practical application in infrastructure asset management.

2. Materials and Methods

2.1. Study Area and Dataset Description

The Kolashin–Podgorica section of the Vrbnica–Bar railway line in Montenegro constitutes a strategically important corridor that traverses mountainous terrain with considerable elevation variation. The alignment incorporates tunnels, bridges, sharp curves, and steep slopes, which create technically demanding conditions for maintaining the stability of track geometry [23] (see Figure 1). These engineering and operational features render this section particularly suitable for the analysis of real-world degradation phenomena.



Figure 1. Location of the Kolashin–Podgorica section on the Vrbanica–Bar railway line (highlighted in red circle). Source: Author’s adaptation based on Railway Infrastructure of Montenegro-AD Podgorica data.

According to geometric data from Railway Infrastructure of Montenegro-AD Podgorica, the Kolashin–Podgorica section extends for 39.7 km, with a minimum curve radius of 300 m and maximum longitudinal slope of 24.9‰ (see Figure 2). More than 65% of the alignment consists of curves, representing one of the highest proportions along the Vrbanica–Bar corridor. The combination of steep slopes s and small-radius curves increases dynamic loading on the track structure, thereby accelerating ballast settlement, rail wear, and geometric instability [24,25].

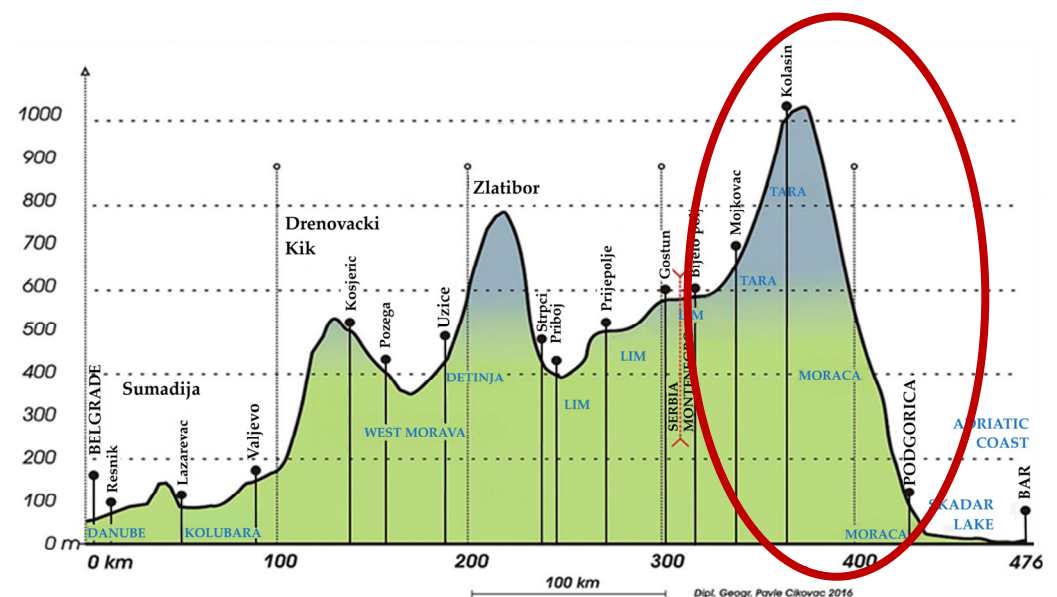


Figure 2. Longitudinal profile of the Belgrade–Bar railway, with the Kolashin–Podgorica section highlighted. Source: Author’s adaptation based on Railway Infrastructure of Montenegro-AD Podgorica data.

These characteristics highlight the considerable operational challenges of this section in comparison to flatter and straighter railway alignments. The high curvature ratio and steep slopes increase the frequency of required maintenance interventions and reduce the long-term effectiveness of tamping operations [26,27]. Consequently, the Kolashin–Podgorica section provides an exemplary case study for predictive modeling of degradation trends under complex geometric and geotechnical conditions.

A structured, multi-step data cleaning workflow was implemented to ensure analytical reliability. The process commenced with the detection and imputation of missing values, followed by the removal of segments with excessive data gaps, the systematic identification of outliers, the verification of outliers against field records, and the correction of anomalous measurement jumps using cross-referenced inspection data.

Track geometry measurements were collected during four independent inspection campaigns in 2017, 2019, 2022, and 2024. Each inspection used a different diagnostic platform per the infrastructure manager’s monitoring schedule: the EM80 vehicle in 2017, the EM120 in 2019, and the Sever 1435 system in 2022 and 2024. All systems operated with a high-resolution spatial sampling interval of 0.25 m, in compliance with European standard EN 13848-1:2019 [28] and UIC Leaflet 518 [29], thereby ensuring statistical comparability and robustness for longitudinal analysis. Measurements were conducted in autumn under weather conditions consistent with those used in previous studies to minimize the effects of seasonal variation. Missing or incomplete measurement values were identified using both automated scripts and manual review and imputed using segment-mean substitution; segments with extensive gaps were excluded from analysis. Outlier detection involved applying a z-score threshold to each geometric parameter within its 20 m segment. Values exceeding three standard deviations from the segment mean were checked against field logs or replaced with the segment median if deemed erroneous. Inconsistent or abrupt measurement jumps not attributable to engineering features were filtered out by cross-referencing inspection records and adjacent-segment data. This comprehensive data-cleaning workflow ensures that the final analytical dataset is both reliable and reproducible. The 2024 data were excluded from model training and calibration and reserved exclusively for final validation, thereby ensuring the objectivity and reliability of the predictive analysis. The main data cleaning steps are summarized as follows: (1) Detection and imputation of missing values, (2) Exclusion of segments with extensive gaps, (3) Outlier identification using z-score thresholding, (4) Field log verification or median substitution for detected outliers, and (5) Removal of erroneous measurement jumps through cross-referencing with inspection records and neighboring data.

2.2. Segmentation into Homogeneous Units

For analytical consistency, the Kolashin–Podgorica section was divided into homogeneous 20-m segments. International standards, such as UIC 518 [28] and EN 13848-1:2019 [29], recommend recording lengths of 10 m, 20 m, and 200 m to balance analytical resolution with operational efficiency. Segmenting at this scale supports both localized anomaly detection and long-term assessment of track fatigue [30].

Spatial abstraction methods in traffic and infrastructure research indicate that homogeneous segmentation supports the integration of heterogeneous datasets and facilitates forecasting of deterioration patterns [31]. Multiscale segmentation improves comparability, enables integration of diverse data sources, and strengthens the link between micro-level defects and macro-level performance metrics [30].

A segmentation length of 20 m represents a practical compromise between precision and stability. Shorter intervals, such as 10 m, increase sensitivity to random fluctuations and measurement noise, often highlighting variations without engineering significance.

In contrast, longer intervals of 200 m obscure localized defects and smooth out critical irregularities, reducing the ability to identify sections requiring urgent maintenance. The 20-m segmentation provides intermediate resolution, ensuring statistical representativeness while maintaining sensitivity to localized degradation zones.

The railway monitoring vehicle collects data at a spatial resolution of 0.25 m, yielding 80 data points per 20 m track segment. This high spatial density establishes a robust statistical basis for calculating standard deviations. It enables effective noise filtering, ensuring that the analyzed geometric parameters accurately reflect actual track conditions rather than sensor artifacts. Future research will investigate multi-scale modeling approaches using longer track segments, 100 to 200 m in length.

This segmentation approach is especially relevant for the Kolashin–Podgorica section, which traverses mountainous terrain with tunnels, bridges, sharp curves, and steep slopes. Finer segmentation is necessary to capture localized effects of infrastructure transitions, such as increased ballast instability at tunnel exits or bridge approaches. Simultaneously, the 20-m interval minimizes within-segment noise and supports stable calculation of statistical indicators.

This segmentation approach enables precise spatial identification of degradation zones [32,33], enhances the reliability of statistical indicators [34,35], and supports model training with sufficient observations to prevent overfitting [36]. Consequently, the 20-m segmentation provides both methodological rigor and practical engineering relevance, facilitating predictive modeling of track quality in geometrically complex railway environments such as the Kolashin–Podgorica corridor.

2.3. Track Quality Index Calculation

Geometric measurements collected by the track recording vehicle at 0.25 m intervals are aggregated into homogeneous 20 m segments. Each segment, containing $n = 80$ discrete data points, is analyzed for the spatial dispersion of each orthogonal track geometry parameter j , where $j \in \{\text{Longitudinal Level, Gauge, Alignment}\}$, using the sample standard deviation (σ_j):

$$\sigma_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{j,i} - \bar{x}_j)^2} \quad (1)$$

In this context, $x_{j,i}$ represents the i -th discrete measurement of parameter j within the segment and $\bar{x}_j$ is the arithmetic mean of that parameter calculated over the same 20 m section. The synthetic local *TQI* for each homogeneous segment is determined by computing the weighted sum of these standard deviations [37]:

$$TQI = \sum_{j=1}^M w_j \times \sigma_j \quad (2)$$

where

- σ_j is the standard deviation of geometry parameter j ,
- w_j is the weight factor assigned based on EN 13848 and UIC 518 standards [28,29],
- M is the number of geometry parameters included in the calculation.

The Combined Standard Deviation (CoSD) method utilizes the standard deviation, a statistical measure of dispersion, to quantify variability across multiple geometric features of railway tracks [38]. Because each geometric parameter affects overall railway performance and safety, aggregating their standard deviations yields an index that is sensitive to both anomalies and gradual degradation.

Weighting factors were established based on engineering relevance and international standards. The selection process involved several steps to ensure methodological rigor and reproducibility. Initially, a systematic review of relevant literature identified recommended weighting schemes from previous studies and international guidelines, particularly EN 13848 and UIC 518 standards [39]. Subsequently, an expert panel of railway engineers and infrastructure specialists assessed the technical significance of each geometry parameter and proposed adjustments to reflect local operational conditions. The proposed weights were then calibrated using empirical degradation data from previous measurement campaigns to verify that the weighting scheme accurately reflected observed trends in track quality. The adopted weights for the TQI components are gauge: 1.20, alignment: 1.00, longitudinal level: 1.50, and twist: 2.00 [27]. This comprehensive approach ensures that the parameter weights summarized in Table 2 are methodologically robust, aligned with best practices for TQI calculation, and provide transparency and reproducibility for future applications. Weighting factors used in the calculation of TQI.

Table 2. Weighting factors used in the calculation of TQI.

Parameter	Weighting Factor (w_i)	Engineering Rationale
Twist	2.0	Highest influence due to its direct effect on vehicle stability
Longitudinal level	1.5	Strong effect on vertical dynamics and ride comfort
Gauge	1.2	Medium influence on wheel–rail contact and wear
Alignment	1.0	Lower influence, but relevant for lateral stability

This method allows for normalization and meaningful integration of parameters expressed in different units (mm, ‰, degrees). Higher TQI values correspond to greater deviations and reduced track quality, whereas lower values indicate well-maintained track conditions. Table 3 presents the interpretation scale:

Table 3. TQI interpretation scale and track condition categories.

TQI Value	Interpretation
0–10	Excellent track quality—very low degradation, no intervention needed.
10–20	Good track quality—low degradation, but preventive inspection may be required.
20–30	Fair track quality—regular maintenance is recommended to prevent deterioration.
30–40	Poor track quality—maintenance is required in the near future to avoid defects.
>40	Critical track quality—urgent repairs are necessary to maintain safety.

The CoSD approach was selected as it provides a single synthetic indicator of track condition that captures both variability and severity across parameters, while remaining interpretable and widely adopted in practice. Compared to alternative methods, such as mean deviation or exceedance-based indices, CoSD offers enhanced comparability across different track sections and inspection periods, which is essential for longitudinal degradation analysis and predictive modeling.

2.4. Input Features

The input parameters for statistical and machine-learning modeling of the TQI are categorized into four groups: geometric parameters, infrastructure characteristics, operational conditions, and maintenance data. This classification enables a comprehensive analysis of the factors influencing track degradation on the Kolashin–Podgorica section.

The selection of variables was based on data availability, engineering relevance, contribution to model accuracy, alignment with established models and recommendations in the academic literature, and compliance with the standards EN 13848 [28] and UIC 518 [29].

A systematic overview of all variables incorporated into the models is provided below.

2.4.1. Track Geometry Parameters

Table 4 presents a systematic classification of the geometric parameters measured and analyzed in this study. For each parameter, an engineering interpretation of its significance and influence on track stability is included, as well as its representation in the model as either a continuous (real-valued) or discrete (binary) variable. These parameters are key indicators of the stability, comfort, and safety of the railway track [40–42].

Table 4. Track geometry parameters, technical significance, and typological categorization. Source: Own data.

Parameter	Description	Type
Track Gauge	Inner distance between the rails, critical for track stability and the safe running of the vehicle on the rails.	Quantitative Continuous
Twist	Slope of the track over short sections—related to ride comfort and the risk of vehicle derailment at increased speeds.	Quantitative Continuous
Alignment (L and R)	Horizontal deviations from the reference line—influence ride stability, cause lateral deviations affecting the wheel–rail contact stability.	Quantitative Continuous
Longitudinal Level (L and R)	Vertical irregularities—influence dynamic loading.	Quantitative Continuous

Note: The geometric parameters (gauge, twist, alignment, and longitudinal profile) are not used individually. Instead, they are aggregated into a single synthetic value using the CoSD method. This process produces a single TQI, which serves as the target variable in the prediction models.

2.4.2. Infrastructure Parameters

Table 5 presents the principal infrastructure characteristics used as input variables in models for analyzing and predicting the TQI. Each parameter is described briefly and classified by data type (binary or continuous). These variables are critical, as they influence the dynamic behavior of the track and contribute to accelerated degradation. Spatial data processing was conducted at the 20-m track-section level to enable detailed analysis of local conditions along the route.

Tunnel presence (0/1): This binary indicator specifies whether a railway segment is located within a tunnel. The confined environment of tunnels typically results in restricted drainage and limited ballast ventilation, both of which accelerate track geometry degradation. Reduced drainage efficiency leads to the accumulation of water and moisture, causing deterioration of the ballast and underlying support, while limited ventilation impedes moisture evaporation. These factors contribute to more rapid degradation of track geometry compared to open track sections. Constrained drainage can result in prolonged exposure

to water-related damage, such as frost action and chemical degradation of ballast materials, directly affecting the long-term integrity of track geometry. Limited airflow in tunnels further exacerbates moisture retention, accelerating material degradation and increasing maintenance demands. These tunnel-specific degradation mechanisms are significant for predictive maintenance models based on track quality indicators, as tunnel sections are frequently associated with a higher risk of accelerated deterioration [43].

Table 5. Route infrastructure parameters, technical significance, and data type. Source: Own data.

Parameter	Description	Type
Tunnel	Section within a tunnel (1/0)—influences drainage, maintenance, and accessibility.	Binary
Bridge	Section on a bridge (1/0)—altered dynamic forces due to rigid connections.	Binary
Railway Station	Section close to/within a station (1/0)—higher stresses due to maneuvering and switches.	Binary
Curve Radius	Measure of route curvature (in m)—below 500 m represents a critical value.	Continuous
Longitudinal Steep slopes	Expressed in ‰—influences contact forces and vertical/horizontal stresses.	Continuous
Sleeper Age	Greater age generally indicates an increased probability of damage and reduced track stability.	Continuous

Note: The steel rails and ballast layer exhibit uniform aging profiles and material homogeneity throughout the continuous alignment, serving as statistical constants. Therefore, sleeper age was selected as the only active superstructure predictor to capture localized structural variance without introducing redundant, zero-variance parameters.

Bridge presence (0/1): This binary indicator specifies whether a track segment is situated on a bridge. Bridges expose the track to elevated dynamic forces, including vibration, impact loads, and differential motion, compared to embankment-supported tracks. These increased forces contribute to localized settlement and impose additional stresses on track components, thereby increasing maintenance requirements. Studies have demonstrated that the dynamic interaction among the train, track, and bridge can result in forces transmitted through bridge structures under high-speed traffic that exceed those experienced on ground-level alignments [44,45].

Station proximity (0/1): This binary indicator reflects whether a track segment is located within or near a station. Literature indicates that such segments are subjected to frequent acceleration and braking forces, which disturb the ballast and create geometric irregularities, thereby increasing maintenance needs. The station proximity indicator is essential in predictive maintenance models, as the dynamic conditions near stations are directly associated with higher degradation rates of ballast and track geometry. Including this indicator in maintenance planning enables railway operators to allocate resources more efficiently and tailor intervention strategies to these high-risk areas [46].

Curve radius (in meters): The horizontal curve radius is a key parameter for evaluating track performance, as smaller radii are associated with increased lateral forces, leading to greater rail wear and accelerated degradation. As the curve radius decreases, trains experience higher centrifugal forces, which increase lateral forces at the wheel–rail interface.

This results in elevated rail wear rates and conditions such as corrugation, contributing to geometric degradation. Studies emphasize that the horizontal curve radius is fundamentally linked to the dynamic loading conditions that drive rail wear and degradation. Precise quantification and management of curve radius effects are therefore vital for predicting maintenance needs and ensuring the long-term reliability of railway infrastructure [47,48].

Longitudinal steep slopes (in ‰): The longitudinal steep slopes, typically expressed in per mille (‰), define the track slope and significantly influence track performance. Steeper slopes require higher traction and braking forces, which impose additional stresses on the ballast and subgrade. These increased forces can intensify substructure settlement, resulting in differential settlement over time. Experimental studies have shown that additional longitudinal forces on sloping tracks are a primary contributor to differential settlement, as increased traction and braking loads heighten dynamic interactions between the train and track [49]. Collectively, these findings indicate that the longitudinal steep slopes are a critical factor affecting the mechanical response of the track substructure. Thorough evaluation of track slope is therefore essential for predictive maintenance models, as it enables proactive assessment of potential settlement issues and informs the development of strategies to mitigate the adverse effects of steep slopes [50,51].

2.4.3. Operational Parameters

Table 6 summarizes the primary operational parameters used as input variables in models that analyze and predict the Track Quality Index (TQI). For the Kolashin–Podgorica section, the principal operational parameters are permissible speed and traffic intensity. These parameters are treated as continuous variables and serve as indirect indicators of mechanical loading and their impact on track stability. Together, they significantly influence the dynamic stability of railway infrastructure, especially under conditions of high train-intensive utilization.

Table 6. Route operational parameters, engineering significance, and data type. Source: Own data.

Parameter	Description	Type
Permissible Speed	Allowed speed on the section (in km/h)—indirect indicator of dynamic loading.	Quantitative Continuous
Traffic Intensity	Total traffic volume in gross tons per section.	Quantitative Continuous

Speed (km/h): The maximum allowable speed for a track segment, measured in kilometers per hour, is established based on infrastructure conditions and regulatory standards. Higher permissible speeds require stricter geometric quality to ensure vehicle stability and safety, as increased dynamic loads are imposed by trains. Track segments designated for high-speed operation must comply with tighter tolerances to maintain the kinematic and dynamic performance necessary for safe railway operation [9,52].

Traffic intensity (gross tons/year): The annual cumulative traffic load, expressed in gross tons per year, is a key indicator of the stress experienced by railway track segments. Increased traffic intensity results in more frequent and intense dynamic forces on the track structure, leading to greater cumulative wear and accelerated geometry degradation. For example, heavy cyclic loading from freight and passenger services intensifies ballast deformation and subgrade settlement, causing track geometry irregularities that require more frequent maintenance [53,54]. Studies evaluating track geometry quality have identified traffic load as a critical factor among other influential variables, demonstrating that segments with higher gross tonnage typically experience greater deterioration in geometric parameters over time [55]. These findings underscore the necessity of incorporating traffic

intensity into degradation-prediction models to improve the accuracy of maintenance forecasting and enhance track infrastructure management.

2.4.4. Maintenance Data

Table 7 presents parameters associated with mechanized track maintenance that serve as input variables for TQI analysis models. These parameters include binary indicators denoting whether track regulation or ballast tamping occurred between measurement campaigns, as well as the total number of interventions. This information enables evaluation of maintenance impacts on track stability and degradation. Each parameter is provided with a concise description and its corresponding data type (binary or continuous).

Table 7. Parameters related to mechanized track maintenance, their significance, and data type. Source: Own data.

Parameter	Description	Type
Mechanized Track Regulation (Tamping)	Whether mechanized regulation of the track geometry was performed in the given period between measurements (1/0). The activity includes lifting and tamping the ballast under the sleepers.	Qualitative Binary
Mechanized Ballast Profiling	Whether mechanized planning and shaping of the ballast prism was performed in the given period between measurements (1/0).	Qualitative Binary
Number of Interventions	Total times of regulation or ballast profiling were performed in the period between two measurement campaigns.	Qualitative Binary

Track tamping/regulation: Track tamping is a primary maintenance intervention for ballasted railway tracks, utilizing specialized machinery to lift and realign the track by inserting vibrating tamping tines into the ballast beneath the sleepers. This process restores both vertical and horizontal alignment, temporarily correcting geometry defects caused by operational loads. While tamping effectively realigns the track, it also disrupts the ballast structure. The vibrations and mechanical actions of the tamping tines can fracture and degrade the ballast, potentially accelerating its deterioration in subsequent loading cycles [56,57]. These observations underscore the engineering trade-offs inherent in tamping as a maintenance strategy. Optimizing tamping to maximize geometry correction while minimizing ballast damage remains a central challenge for railway infrastructure management. Ongoing field investigations and simulation studies provide essential insights for refining operational parameters and scheduling tamping interventions to ensure long-term track performance and safety [9,58,59].

Ballast profiling and shaping: This mechanized maintenance activity is essential for restoring the geometry of ballast shoulders and crib spaces, which is necessary to maintain the lateral stability of sleepers and enhance drainage performance. Dedicated machines, such as dynamic track stabilizers, are typically deployed after tamping operations to reestablish the optimal ballast profile [60]. Ballast profiling and shaping are fundamental for reconfiguring the ballast layer, thereby supporting sleeper stability and improving drainage, both of which are critical for mitigating long-term degradation of railway infrastructure.

Counts per period: The total number of interventions, including both tamping/regulation and ballast profiling/shaping, conducted during the three analysis intervals

(2017–2019, 2019–2022), serves as a key metric. This measure links the frequency of restorative actions to observed variations in track degradation rates [61]. Maintaining accurate records of intervention counts over defined periods enables quantification of infrastructure workload and assessment of maintenance strategy effectiveness. Analyzing maintenance counts alongside degradation trends over consistent intervals provides a clearer understanding of the relationship between scheduled interventions and the long-term evolution of track geometry quality.

To explicitly capture the influence of track maintenance on predictive model behavior, discrete frequencies of maintenance interventions—specifically, the number of localized tamping activities in each 20-m segment—were incorporated into the model’s input feature matrix X as independent predictors. An audit of empirical records for the 2022 to 2024 evaluation period indicates a complete absence of systematic, corridor-wide mechanized tamping overhauls. Operational logs show that most 20 m segments experienced zero or one isolated localized intervention, primarily due to budgetary constraints. Consequently, the track layout did not undergo large-scale geometry resets by heavy machinery during this period, and the independent TQI2024 target dataset reflects a continuous, natural progression of degradation. This operational context supports the validity of the modeling framework without requiring structural adjustments to account for extensive maintenance events.

The limited number of maintenance events during the 2022 to 2024 period may restrict the generalizability of the model’s findings. The dataset underrepresents scenarios involving more frequent or systematic maintenance, which may reduce the robustness of the relationships learned between maintenance activity and degradation rates when applied to other railways or periods with different maintenance regimes. This limitation should be considered when interpreting the results and evaluating the broader applicability of the predictive models developed in this study. To address this constraint, future studies or practical applications should expand data collection to include a wider range of maintenance activity levels, both in frequency and type, by extending the observation window or targeting lines with differing maintenance regimes. Alternatively, synthetic data simulations or scenario modeling could be employed to examine how increased or varied maintenance activities affect degradation trends and model outcomes. These approaches would enhance the robustness and generalizability of predictive models across diverse operational contexts.

To address this limitation, future studies should incorporate data from railway lines or periods with a broader range of maintenance activities, including both frequent and systematic interventions. This objective can be achieved by extending the observation window, collaborating with infrastructure managers across different networks, or integrating external datasets to capture diverse maintenance regimes. Additionally, simulating various maintenance scenarios or augmenting datasets with synthetic data could provide valuable insights into the effects of different maintenance strategies. These approaches would improve model robustness and generalizability across diverse operational and maintenance contexts.

From a decision-making perspective, model outputs from this dataset are most reliable for corridors or periods with similarly low maintenance activity. For networks with more intensive or regular maintenance schedules, predictive reliability regarding the effects of maintenance actions may be reduced. Therefore, planning and resource allocation decisions should account for this limitation and, where possible, incorporate additional data representing a broader range of maintenance scenarios to supplement or recalibrate the models for other contexts. To adapt the predictive framework for rail lines with different maintenance regimes, a structured adaptation process should be followed. This

process involves collecting and integrating data on maintenance actions that reflect the specific characteristics of the new environment, including the frequency, type, and timing of interventions. Where feasible, the model should be recalibrated by retraining on local data encompassing a broader spectrum of maintenance activities.

Additionally, model performance should be validated against recent periods that include both typical and intensive maintenance episodes, with adjustments to feature selection or input representations as necessary. Regular local revalidation and calibration will help ensure the model remains robust and accurately represents the maintenance-degradation relationship in varied operations. To address this limitation and improve model robustness, future research should expand the dataset to include periods with more frequent or varied maintenance activities. Supplementing the current data with external datasets from other railway lines or collaborating with infrastructure managers to capture a broader range of maintenance regimes would enhance the model's generalizability. Furthermore, integrating data from adjacent railway sections or applying synthetic data augmentation techniques to simulate a wider array of maintenance scenarios would enable a more accurate assessment of maintenance impacts on track degradation and support the development of more resilient predictive maintenance strategies.

Track geometry variables are directly incorporated into the TQI calculation in accordance with EN 13848 and UIC 518, while infrastructure, operational, and maintenance parameters serve as boundary conditions that indirectly affect degradation processes. This integrated feature set enables predictive models to capture both the intrinsic geometric condition of the track and the external factors influencing its evolution over time. Numerical features were scaled using MinMaxScaler to ensure consistent ranges and eliminate scale bias across variables. Explicit feature selection or dimensionality reduction techniques, such as principal component analysis (PCA), were not applied; instead, all relevant variables identified in previous sections were retained in the modeling pipeline to preserve interpretability and enable assessment of each input's contribution to predictive performance. Feature selection was intentionally omitted to maximize transparency and interpretability of model outputs, facilitating analysis of the influence of each engineering and operational input on track degradation. While excluding variables or applying dimensionality reduction could have improved computational efficiency and reduced model complexity, this would have diminished insight into the role of individual predictors. Therefore, interpretability for practical maintenance assessment and engineering understanding was prioritized over marginal gains in predictive performance. Nevertheless, alternative approaches, such as applying regularization methods or PCA to reduce dimensionality, could be explored in future studies to further balance interpretability with predictive accuracy and computational efficiency.

2.5. Modeling Approach

This subsection outlines the modeling framework utilized in the study, including the processes of model selection, training methodology, and validation procedures.

The predictive pipeline operates through distinct computational phases designed to systematically map patterns of track quality degradation. As shown in Figure 3, the integrated architecture incorporates a comprehensive set of input parameters, such as infrastructure boundary layouts (bridges, tunnels, stations), macro-alignment geometry (horizontal curves and longitudinal steep slopes), operational stress indicators (permissible line speed and traffic tonnage), and cumulative maintenance records, to establish the core modeling baseline. The temporal framework is strictly partitioned to avoid any data leakage: the historical dataset spanning 2017 to October 2022 is utilized exclusively for feature engineering, model training, hyperparameter tuning, and cross-validation. Con-

versely, the 2024 campaign profile is strictly reserved as the final, independent test set for evaluating forward-predictive capacity, ensuring that no post-2022 operational or structural data influences the model training loops.

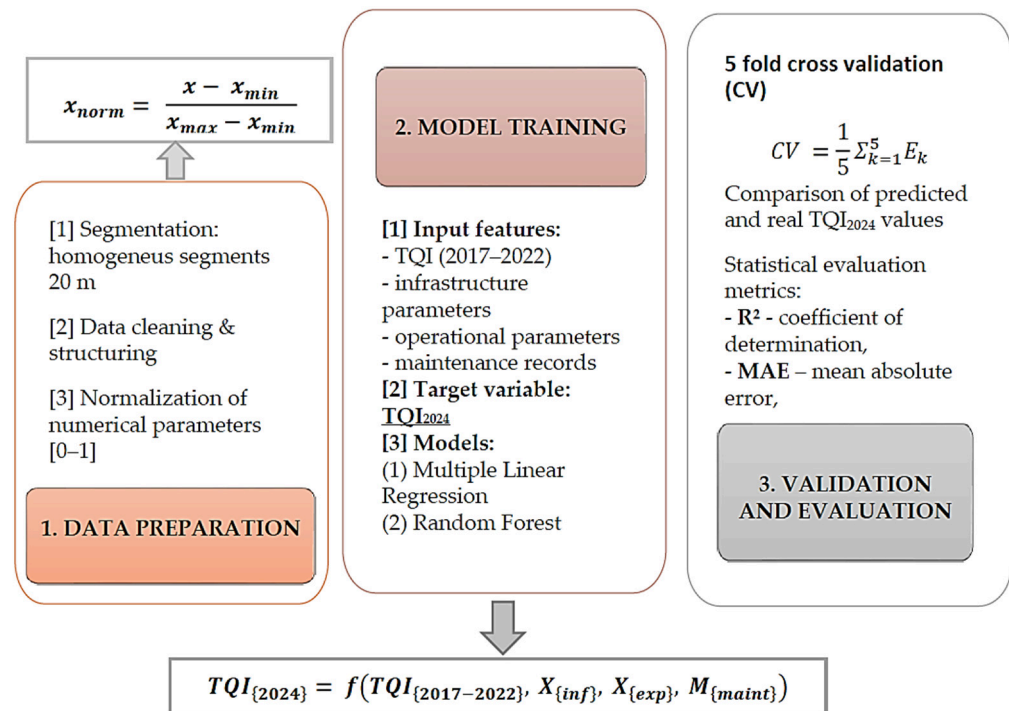


Figure 3. Methodological framework of the predictive pipeline highlighting the input features, historical training epochs (2017–2022), and the isolated independent validation target variable (TQI_{2024}) (Author’s original illustration based on the research methodology).

A comprehensive normalization pipeline was implemented to ensure numerical consistency across heterogeneous input vectors and eliminate scaling discrepancies. Statistical and machine learning models are sensitive to variations in feature magnitude, as variables with larger absolute values can disproportionately influence cost functions during optimization [62,63]. Consequently, the standard Min-Max scaling method was applied, transforming each numerical feature into a bounded [0, 1] range according to the following mathematical formulation:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

where x represents the actual empirical observation, while x_{min} and x_{max} denote the minimum and maximum boundaries of that specific feature across the entire dataset.

Raw track geometry deviations include algebraic signs to represent spatial directions, such as negative and positive values for left or right alignment shifts and gauge widening or narrowing. The data preprocessing pipeline transforms these parameters using their absolute magnitudes or scales them relative to specific directional tolerance limits prior to normalization. This approach prevents counter-directional variations, such as alternating positive and negative twist peaks, from canceling each other mathematically and preserving essential spatial volatility information.

This non-negative linear transformation preserves the underlying distribution structure and ensures that all engineering variables, such as maximum permissible speed, annual traffic tonnage, alignment profiles, and longitudinal steep slopes s , contribute symmetrically to the modeling process. The target vector (TQI_{2024}) was also scaled during training to maintain steep slope stability and subsequently back-transformed to its original engineering scale for realistic performance evaluation and direct physical interpretation.

This methodology supports stable model training and robust distance metrics, preventing numerical skewing during matrix optimization phases [64].

A sequential algorithmic blueprint was established to ensure rigorous execution of the analytical cycle and eliminate potential data inconsistencies. Figure 3 delineates the modeling pipeline across three foundational phases: data preprocessing and stabilization (Phase 1), machine learning and statistical baseline configuration (Phase 2), and objective validation with statistical performance metric extraction (Phase 3). During Phase 1, all feature matrices containing spatial information from the 2024 campaign are fully isolated from the optimization loops to eliminate data leakage and prevent future infrastructure profiles from influencing training weights. Additionally, continuous Min-Max normalization is applied to all continuous features to mitigate magnitude dominance due to heterogeneous feature boundaries, resulting in a stabilized, dimensionless $[0, 1]$ range.

Two supervised regression models were implemented using Python's scikit-learn library. The modeling objective was to predict the TQI value for each 20-m segment in the 2024 inspection, utilizing features derived from the 2017–2022 period. These models were selected to balance interpretability and predictive performance. Multiple Linear Regression (MLR) provides a transparent and interpretable baseline for understanding the influence of individual input factors, while the Random Forest Regressor (RF) enhances predictive accuracy by capturing complex, non-linear relationships in the data. This combination enables both the explanation of model outputs and robust forecasting results. The models used were:

- Multiple Linear Regression (MLR) and
- Random Forest Regressor (RF)

2.5.1. Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) is a fundamental parametric statistical technique used to model the relationship between a single continuous dependent variable and multiple independent input variables [65,66]. In this analysis, MLR is employed as the baseline statistical model to forecast the track quality index (TQI_{2024}) using multi-epoch historical parameters from 2017 to 2022. The primary objective is to determine whether a linear relationship exists between integrated geometric, infrastructural, operational, and maintenance features and the future geometric state of the track structure.

The MLR framework assumes a linear dependency between the target vector Y and the independent predictors $X_1, X_2, \dots, X_m$. The general mathematical formulation of the model is presented below:

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i + \sum_{i=1}^n \beta_{ii} X_i^2 + \sum_{i < j} \beta_{ij} X_i X_j + \varepsilon \quad (4)$$

where

- Y is the target predicted variable (TQI_{2024}),
- X_i represents the continuous or binary standardized independent input features,
- β_i denotes the partial regression coefficients estimated via the regression analysis, representing the sensitivity of the target to each predictor,
- ε is the random residual error term of the model, assuming an expected mean of $E(\varepsilon) = 0$ and a constant variance of σ^2 ($\varepsilon \sim N(0, \sigma^2)$).

The parameter optimization process aims to determine the β_i coefficients that minimize the residual error objective function. This optimization is conducted using the ordinary

least squares (OLS) method, which minimizes the sum of squared differences between actual and predicted values [65,66]:

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(Y_i - \left(\beta_0 + \sum_{j=1}^m \beta_j X_{j,i} + \sum_{j=1}^m \beta_{jj} X_{j,i}^2 \right) \right)^2 \quad (5)$$

As a foundational benchmark, MLR provides a transparent and interpretable framework for capturing the direct sensitivities of the engineering state of the permanent way to variations in external loading. Although it simplifies the underlying physical degradation mechanisms by assuming strict linearity, its application is essential for establishing baseline performance, verifying data consistency, and evaluating the necessity for more advanced machine learning architectures.

2.5.2. Random Forest Regressor (RF)

The Random Forest (RF) Regressor is a non-parametric ensemble learning method developed for high-dimensional supervised regression tasks. This method constructs an ensemble of decentralized, uncorrelated decision trees, collectively termed a ‘forest,’ to generate a joint prognostic response that is more robust, stable, and less susceptible to variance inflation than any individual decision tree [67].

The RF algorithm operates according to three interconnected structural principles:

- **Bootstrap Aggregation (Bagging):** Given a historical training dataset D of size $N = 821$ (representing the 20-m homogeneous track segments from 2017 to 2022), the algorithm generates B distinct bootstrap samples D_b (where $B = n_{\text{estimators}} = 100$). Each subset of D_b is constructed by randomly sampling N observations with replacement from D [68]. Consequently, some track segments may appear multiple times within a single bootstrap sample, while approximately 36.8% of the observations are withheld as Out-of-Bag (OOB) data. This stochastic partitioning introduces localized variation among the individual decision trees and reduces the model’s susceptibility to overfitting.
- **Stochastic Feature Subspace Selection:** During recursive binary splitting within each tree, the algorithm evaluates a randomized subspace of m predictors at each node (typically $m = M/3$). The optimal split is identified by maximizing feature purity or minimizing residual impurity, such as Mean Squared Error (MSE), within this subset [69]. This approach decorrelates the trees and prevents highly correlated parameters from dominating each split.
- **Mathematical Ensemble Aggregation:** For regression frameworks, the final predicted Track Quality Index ($TQ\hat{I}_{2024}$) for an unseen 20-m evaluation segment x is computed as the arithmetic mean of the outputs from all B independent bootstrapped decision trees [68]:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (6)$$

Here, $f_b(x)$ represents the regression output of the b -th individual decision tree. This averaging mechanism reduces variance and stabilizes the index output against localized anomalies.

After preliminary hyperparameter optimization using grid-search multi-fold cross-validation, the RF framework was configured to balance generalization capacity, robustness to input sensitivity, and computational complexity:

- **Number of Trees ($n_{\text{estimators}} = 100$):** This value ensures statistical convergence of the ensemble average. Increasing the tree count beyond 100 for an 821-row dataset

provides diminishing returns, as prediction variance stabilizes and the model becomes insensitive to random feature initialization.

- Maximum Tree Depth (max_depth = 8): Limiting tree depth to 8 acts as a regularization strategy, preventing overfitting to localized measurement noise and supporting robust predictive performance across track sections.
- Minimum Samples Split (min_samples_split = 2 or 5) and Features (max_features = "sqrt"): These settings decorrelate individual trees, reduce sensitivity to dominant predictors, and stabilize the model.
- Reproducibility Anchor (random_state = 42): This parameter ensures consistent pseudo-random number generation, supporting full experimental reproducibility under identical data pipelines.

The robustness of this hyperparameter configuration is validated by Out-of-Bag (OOB) error monitoring, which provides continuous internal validation and demonstrates insensitivity to parameter fluctuations on unseen tracking data. Although Random Forest is not sensitive to feature scaling, MinMax scaling was applied to all numerical inputs to ensure fairness, consistency across models, and a unified analytical pipeline. After training, predictions for TQI₂₀₂₄ were obtained and then inverse-transformed to the original scale for evaluation and comparison. A 5-fold cross-validation routine was implemented to ensure reliable, generalizable, and unbiased evaluation of the predictive models. Although inherent spatial autocorrelation exists in sequential 20 m segments, this approach mitigated over-optimistic validation metrics during training. Final models were evaluated against an independent, chronologically isolated 2024 dataset to objectively assess generalization. As illustrated in Figure 4, the historical training data (2017–2022) was partitioned into five non-overlapping blocks for sequential iterations. In each iteration, four folds were used for model optimization, while the remaining fold served as an internal validation benchmark. To explicitly address spatial autocorrelation along the continuous linear corridor, an additional chainage-blocked cross-validation strategy (GroupKFold grouped by spatial sections) was implemented alongside the standard cross-validation routine. This ensures that training and validation folds do not share adjacent, strongly correlated 20 m segments, providing a rigorous test of the model's spatial generalization and sensitivity. GroupKFold was chosen because it explicitly partitions the dataset based on predefined spatial groups, ensuring that training and validation folds do not share adjacent, strongly correlated 20 m segments. This provides a rigorous test of the model's spatial generalization and sensitivity, and minimizes random or structural bias, resulting in a more robust baseline prior to final deployment. As illustrated in Figure 4, the historical training data (2017–2022) was partitioned into five non-overlapping blocks for sequential iterations. In each iteration, four folds were used for model optimization, while the remaining fold served as an internal validation benchmark. This structure minimizes random or structural bias and provides a robust baseline prior to final deployment.

This unified framework enables direct comparison of results and provides a reliable basis for evaluating the suitability of statistical and machine learning approaches for predictive railway maintenance planning.

For practical application, the following implementation steps provide a general framework. These steps should be adapted to the specific conditions, data availability, resources, and operational requirements of each railway or organization.

1. Pilot the framework on a selected railway corridor by collecting and processing high-resolution track geometry and infrastructure data as described in this
2. Apply the outlined data cleaning, segmentation, and normalization procedures to prepare datasets in accordance with established standards (EN 13848 and UIC 518) [28,29], ensuring methodological consistency and comparability.

3. Select relevant input features based on local infrastructure specifics and integrate operational and maintenance records to tailor the predictive models to the unique context of the line under study.
4. Apply the described modeling pipeline, utilizing both MLR and RF as reference approaches, and conduct rigorous validation using a temporally isolated dataset to assess predictive performance.
5. Interpret model outputs with attention to variable importance and identified degradation hotspots. Use these insights to inform targeted maintenance planning, adjust inspection intervals, and optimize resource allocation for preventive interventions.
6. Adapt and iterate the framework by incorporating additional features or alternative models based on feedback from field deployment, monitoring improvements in forecasting accuracy, and operational outcomes.



Figure 4. Visual schema of the 5-fold cross-validation procedural framework across the historical dataset iterations. (Author’s original illustration based on the research design).

Integrating this predictive modeling approach into routine maintenance planning requires collaboration between engineering and data analysis personnel and access to appropriate computational resources. Implementation requires track geometry and infrastructure datasets (see Section 2.4), standard computing equipment or cloud-based platforms capable of running Python scripts within the Google Colab environment (Python 3.10), and staff with expertise in data analysis or railway engineering. Core software components, including Python and open-source libraries such as scikit-learn, pandas, and matplotlib, are freely available, minimizing software-related barriers. Typically, a small team consisting of an engineering project lead and a data analyst is sufficient for deployment and ongoing operation.

Key implementation steps for deploying the predictive model in practice are as follows:

1. Prepare and collect high-resolution track geometry, infrastructure, operational, and maintenance data for the selected railway corridor.
2. Clean and preprocess the dataset according to established standards, including missing value imputation, outlier removal, and 20-m segmentation.
3. Normalize numerical features for consistency and compile all relevant input variables into a unified feature matrix as specified in the methodology.
4. Set up the computational environment using Python with required open-source libraries (such as scikit-learn, pandas, and matplotlib) on local computers or cloud-based platforms.
5. Train and calibrate both the MLR and RF models using historical inspection data, following the described pipeline and cross-validation strategy.

6. Validate model performance using an independent set of recent inspection data to ensure unbiased assessment of predictive accuracy.
7. Interpret model outputs, focusing on site-specific risk indicators and segment-level predictions to prioritize maintenance actions.
8. Integrate the predictive outputs into existing asset management or maintenance planning workflows to inform decision-making.
9. Periodically update and recalibrate the models by incorporating new inspection and maintenance records as they become available.

Systematic adherence to these steps enables the framework to be scaled and adapted to diverse railway environments, supporting proactive track asset management and enhancing the reliability of maintenance decision-making.

2.6. Software Tools and Computational Infrastructure

All modeling and evaluation tasks were performed using the Python programming environment, which is a widely adopted platform for statistical analysis and machine learning. The following libraries and frameworks facilitated these tasks:

- scikit-learn: utilized for implementing regression models, including MLR and RF, and for conducting cross-validation procedures. This library provided a standardized framework for defining, training, and validating predictive models [70,71].
- statsmodels: used for detailed parameter diagnostics and statistical validation of regression-based approaches, including confidence intervals, hypothesis testing, and residual diagnostics [70,72].
- pandas and numpy: applied for dataset integration, data transformation, and numerical computations [72].
- matplotlib: used to visualize results, including error distributions, scatter plots, and comparative histograms [70].
- Google Colab: functioned as the computational environment, providing GPU acceleration and supporting reproducibility through cloud-based scripting. Direct integration with Google Drive facilitated organized storage and collaboration [72].

The adoption of these open-source tools ensured methodological transparency, computational efficiency, and full reproducibility of the study. This approach aligns the research design with current best practices in infrastructure data analysis and predictive modeling.

2.7. Evaluation Metrics

Model accuracy and error were assessed using the following evaluation metrics:

- R^2 (Coefficient of Determination)

R^2 quantifies how closely predicted values align with actual target values. Higher R^2 values indicate superior model performance.

$$R^2 = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} \quad (7)$$

In this context, y_i denotes the observed values, $\hat{y}_i$ represents the predicted values, and $\bar{y}$ is the mean of the observed values. For track quality assessment, a higher R^2 demonstrates the model's capacity to accurately reproduce degradation trends [73,74].

- MAE (Mean Absolute Error):

MAE measures the average magnitude of prediction errors, regardless of their direction.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

Within railway applications, a lower *MAE* indicates that deviations between predicted and actual TQI values are minimal on average, which reduces the risk of misclassifying track segments by quality [75].

- RMSE (Root Mean Square Error)

RMSE places greater emphasis on larger errors compared to *MAE*, thus offering a more sensitive measure of prediction precision.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

Because extreme prediction errors can lead to underestimating the severity of critically degraded segments, *RMSE* is essential for evaluating reliability in maintenance planning. These metrics were chosen due to their widespread adoption in statistical and machine learning studies on infrastructure condition modeling, which ensures comparability with existing literature [76]. Collectively, they provide a balanced assessment of explanatory power and predictive accuracy, both of which are critical in railway infrastructure applications.

For practical decision-making, it is essential to interpret these metrics within the appropriate operational context:

- R^2 values above 0.70 generally indicate a strong model fit. Values between 0.50 and 0.70 are considered moderate and remain suitable for trend analysis. Values below 0.50 suggest limited predictive relevance and may require further model refinement.
- For *MAE* and *RMSE*, lower values indicate higher predictive accuracy. As a reference, *MAE* and *RMSE* scores that are substantially smaller than the typical range of TQI values observed in routine assessments are generally acceptable. However, acceptable error thresholds should align with operational requirements and the railway manager's risk tolerance. For instance, an *MAE* below 10–15% of the mean TQI may be considered satisfactory, although this benchmark can vary depending on the intended application of the predictions.
- Additionally, when comparing model results, a small difference between *RMSE* and *MAE* indicates that the model is not significantly influenced by large outliers or extreme errors.

These interpretation guidelines enable managers and engineers to assess model outputs more effectively, thereby supporting informed decisions about maintenance priorities and resource allocation.

3. Results

Empirical and quantitative results from predictive modeling of TQI degradation on the Kolashin–Podgorica railway section are presented. The analysis commences with a detailed assessment of variable contributions using Random Forest-based feature importance metrics, thereby establishing a robust, data-driven foundation. Key infrastructure and operational stressors are identified to provide a contextual framework for subsequent predictive modeling. The comparative performance of the baseline MLR and advanced RF models is evaluated. Model development and predictive capabilities are assessed using a

5-fold cross-validation strategy on the historical dataset (2017–2022) and further validated on the independent 2024 dataset.

The findings offer asset managers actionable strategies to enhance maintenance planning and resource allocation. Model outputs enable precise identification of high-risk segments, supporting targeted scheduling of interventions such as tamping or renewal in areas most susceptible to rapid deterioration. This proactive approach facilitates more efficient deployment of maintenance teams and extends the service life of track assets. Integration of predictive risk indicators into asset management systems enables early warning and prioritization for segments exhibiting early signs of accelerated degradation. These data-driven strategies support optimal adjustment of maintenance schedules and resource allocation, thereby improving cost efficiency and operational reliability throughout the corridor.

3.1. Quantitative Assessment of Feature Importance

A comprehensive quantitative assessment of aggregated feature importance from 2017 to 2024 was performed using the RF framework to advance engineering and mathematical understanding of track-asset deterioration along the geotechnically complex Kolashin–Podgorica corridor (see Figure 5). The findings indicate a clear predominance of macro-morphological track geometry parameters. In particular, the horizontal curve radius (R) is identified as the most influential factor, with an aggregated importance of 38.5%, followed by longitudinal steep slopes s (slope) at 31.3%. This distribution highlights significant dynamic lateral forces and vertical load amplifications along this Balkan corridor, where sharp curvature profiles (R between 300 m and 500 m) often coincide with steep longitudinal slopes of up to 25‰ at an operational speed of 60 km/h.

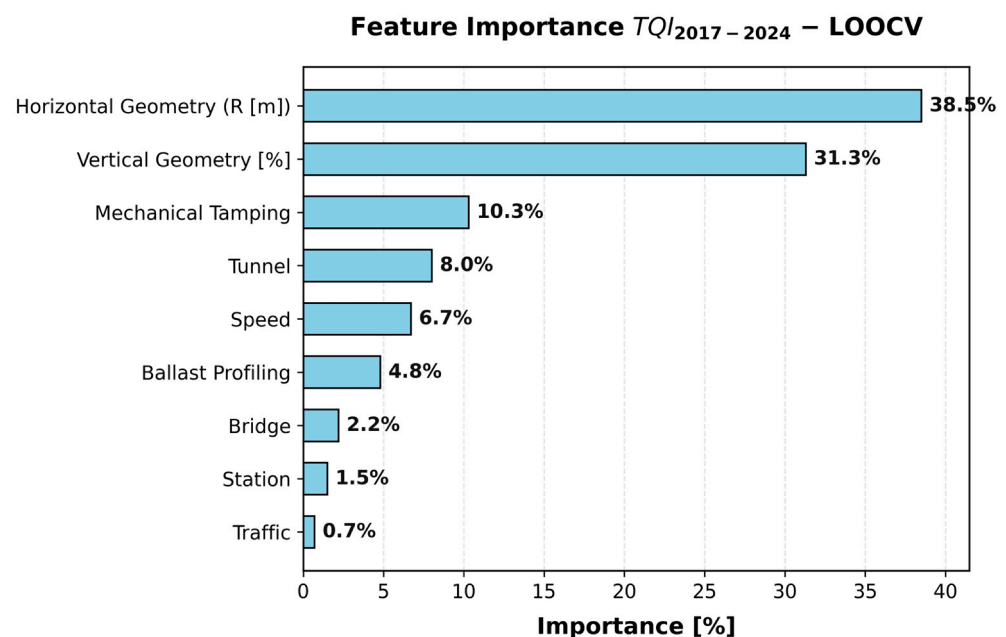


Figure 5. Aggregated feature importance scores for the composite TQI predictive framework across the complete multi-epoch analytical window (2017–2024), computed using RF recursive partitioning combined with Leave-One-Out Cross-Validation (RF-LOOCV). (Source: Author’s calculations based on model outputs).

The model quantifies the influence of localized maintenance interventions, such as tamping and profiling, which together account for approximately 16% of the overall importance. This effect becomes more pronounced during the 2022–2024 period, when

maintenance activity reached its highest levels, with 1.56 leveling and 0.94 profiling actions per 20-m homogeneous segment, indicating increased structural volatility.

The analysis further elucidates the overlapping, non-linear cross-correlations between spatial features and geometric boundaries. Static structural assets, such as tunnels, exhibit stable and increasing standalone importance, reaching 10% by 2022/2024, due to accessibility constraints, microclimate moisture accumulation, and inadequate drainage. In contrast, other fixed elements, including bridges (3.53% of the alignment) and railway stations (5.36%), demonstrate minimal direct feature importance, remaining below 5%. Gross traffic tonnage exerts negligible statistical influence due to its spatial uniformity along the corridor. However, when assessed individually, these infrastructure parameters display considerable overlap in their effects. Their structural impact on TQI degradation is primarily indirect, intensifying local track stiffness discrepancies at rigid boundary transition zones, such as bridge approaches and tunnel portals, where sharp curves and steep slopes converge. The ensemble Random Forest architecture, through recursive partitioning, effectively delineates these multi-variable overlapping clusters, ensuring that macro-averaged line indicators do not obscure localized structural anomalies.

3.2. Predictive Performance of MLR and RF

Model training, hyperparameter tuning, and cross-validation were conducted exclusively on the historical dataset spanning 2017 to October 2022. To assess true forward-predictive capacity, the final evaluation was conducted solely on the independent 2024 dataset, with no post-2022 data utilized during feature engineering or model training.

To rigorously quantify improvements in predictive performance and confirm the statistical significance of the proposed RF model, a traditional engineering benchmark was established. The Naive Persistence Baseline (TQI_{pers}) assumes static track geometry conditions throughout the forecasting horizon, setting the predicted 2024 value equal to the empirical baseline measured in 2022:

$$TQI_{pers,i}(2024) = TQI_i(2022)$$

Here $TQI_{pers,i}(2024)$ represents the baseline prediction for track segment i . This baseline prediction vector was evaluated against the true empirical ground truth, as compiled in Table 8.

Table 8. Final independent model evaluation metrics for the predicted 2024 Track Quality Index (TQI_{2024}). (Source: Author’s calculations based on model outputs).

Model	R^2	MAE	RMSE
Linear Regression	0.5696	4.6805	6.3013
Random Forest	0.6895	3.9630	5.3518
Naive Persistence Baseline (TQI_{2022})	0.2047	5.8634	8.5653

The independent 2024 evaluation results demonstrate a consistent performance trend, with the RF model achieving higher accuracy and lower structural errors than both the Naive Persistence baseline and the MLR model. On the independent 2024 target dataset, RF achieved an R^2 of 0.6895, an MAE of 3.9630, and an RMSE of 5.3518. Random Forest reduced the average prediction error on the unseen 2024 data by approximately 15.3% (MAE) and 15.1% (RMSE) relative to the MLR model and achieved a 32.4% reduction in MAE and a 37.5% reduction in RMSE compared to the Naive Persistence benchmark. Paired permutation tests on out-of-sample residuals between RF and MLR predictions

yielded p -values below 0.01 for both MAE and RMSE, confirming that the performance improvements achieved by RF over MLR are statistically significant. These improvements have important practical implications for engineering applications. Higher predictive accuracy enables maintenance planners to more reliably identify track segments at greatest risk of deterioration, allowing for more precise targeting of interventions and more efficient use of limited resources. Lower MAE and RMSE values lead to fewer unexpected errors in predicted track quality, reducing the likelihood of missing critical defects or of performing unnecessary maintenance on healthy segments. By linking these statistical results to operational decisions, railway managers can improve the planning and prioritization of tamping, renewal, and other interventions, ultimately increasing the safety, reliability, and cost-effectiveness of maintenance operations.

Improvements in statistical error metrics yield tangible operational benefits for railway maintenance. Reduced mean absolute error ensures that model forecasts for individual track segments align more closely with actual conditions, reducing the risk of overlooking deteriorating sections or misallocating maintenance resources. Enhanced predictive precision enables engineers to proactively identify and address high-risk zones before defects escalate to service-affecting failures, thereby minimizing emergency interventions and supporting more cost-effective, scheduled maintenance. These accuracy gains can extend asset life, improve service reliability, and optimize budgeting by reducing reactive spending and targeting interventions for maximum impact.

To further address potential spatial autocorrelation and meet rigorous cross-validation standards for continuous linear infrastructure, an additional chainage-blocked cross-validation strategy (GroupKFold, grouped by spatial sections) was implemented. As expected, evaluating the models on entirely unseen geographical blocks yielded more conservative metrics, achieving an R^2 of 0.4111 and an RMSE of 7.3703 for the Random Forest model, and an R^2 of 0.2658 and an RMSE of 8.2299 for Linear Regression. While these blocked validation scores are lower due to the removal of spatial interpolation bias, they confirm the model's underlying stability and generalizability across different segments of the corridor.

It is important to explicitly state the key modeling assumptions underlying both the MLR and RF approaches to allow critical assessment of the results. First, both models assume the stationarity of relationships between input features and the TQI across the time window and along the studied corridor; that is, the statistical properties and relationships among variables are presumed not to change significantly over the study period. Second, there is an implicit assumption of feature independence, particularly in linear regression, which may not hold due to potential correlations among infrastructure and operational parameters. Third, the models rely on the completeness and accuracy of the selected input features, presuming that the most influential variables affecting track degradation are adequately captured. For MLR, linearity between predictors and the target variable is an explicit assumption.

In contrast, for RF, the method allows for nonlinear relationships and interactions but still assumes that training data sufficiently represent the range of future scenarios. Finally, both approaches assume that the spatial segmentation and preprocessing steps effectively address measurement inconsistencies and that the training and test data are drawn from the same underlying distribution. Violation of any of these assumptions, such as significant changes in operational practices, infrastructure upgrades, or the influence of unmeasured external factors, could affect model performance and the transferability of results to other contexts or time periods.

These considerations highlight the need for cautious interpretation of the results and ongoing model validation as new data becomes available.

These statistically significant improvements directly enhance maintenance planning by enabling more accurate identification of track segments at higher risk of deterioration, prioritizing interventions, and allocating resources to areas of greatest need. Enhanced predictive precision supports earlier detection of emerging defects, reducing the likelihood of unexpected failures and emergency repairs. As a result, engineers can make better-informed decisions, lowering operational risk and supporting cost-effective, proactive maintenance strategies.

These results demonstrate that track degradation is not a simple, isolated function of time but is instead highly non-linear and driven by a complex interplay of spatial and operational factors, including structural boundary conditions such as tunnels and bridges, geometric alignment such as curves, and dynamic traffic loading. By mapping these multivariate interactions, the ensemble machine learning architecture effectively captures localized variations in deterioration that standard historical extrapolation methods cannot address.

3.3. Error Distribution and Visual Analysis

Graphical comparisons provide additional insight into the predictive performance of the two models, supplementing the information obtained from aggregated metrics.

Figure 6 presents a scatter plot of actual versus predicted TQI_{2024} values generated by the Linear Regression model. The red line ($y = x$) represents the ideal fit, and the distance of each point from this line indicates the magnitude of prediction error. Most data points are concentrated in the lower TQI range ($TQI < 20$). As TQI values increase, dispersion around the diagonal becomes more pronounced. For segments with TQI greater than 30, the trend line (black dashed) in Figure 6 deviates substantially from the ideal fit (red line).

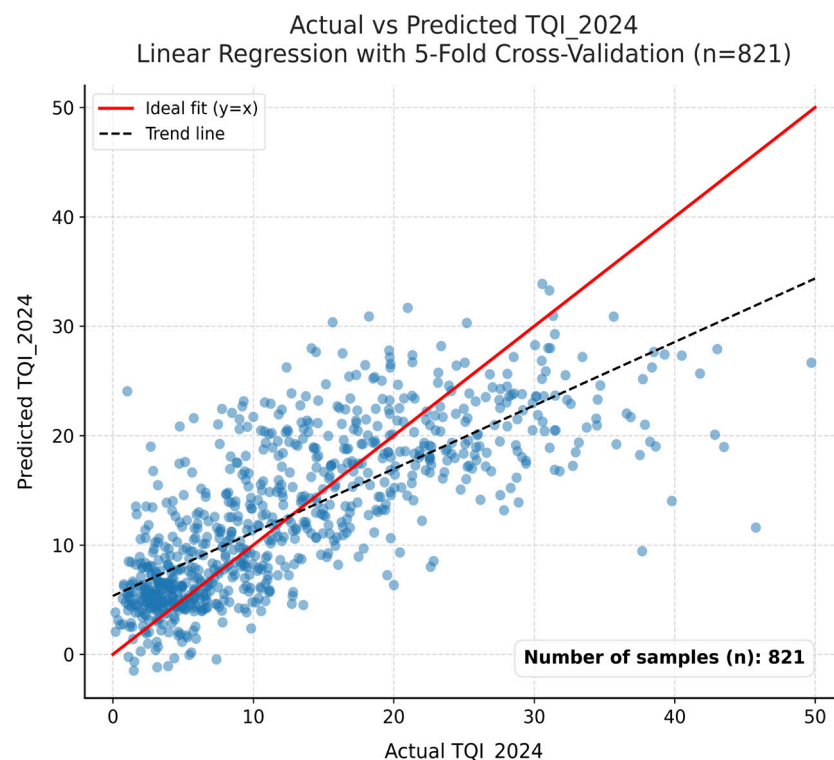


Figure 6. Actual vs. predicted TQI_{2024} values using Linear Regression with 5-fold cross-validation, with the red line ($y = x$) indicating the ideal fit. (Source: Author's calculations based on model outputs).

Figure 7 presents a histogram comparing the actual and predicted distributions for the Linear Regression model. Within the TQI range of 0–20, the predicted frequency closely

matches the actual distribution. For higher TQI values ($TQI > 25$), the model predicts lower frequencies than observed, and local deviations are evident in the intermediate range ($TQI = 10\text{--}15$).

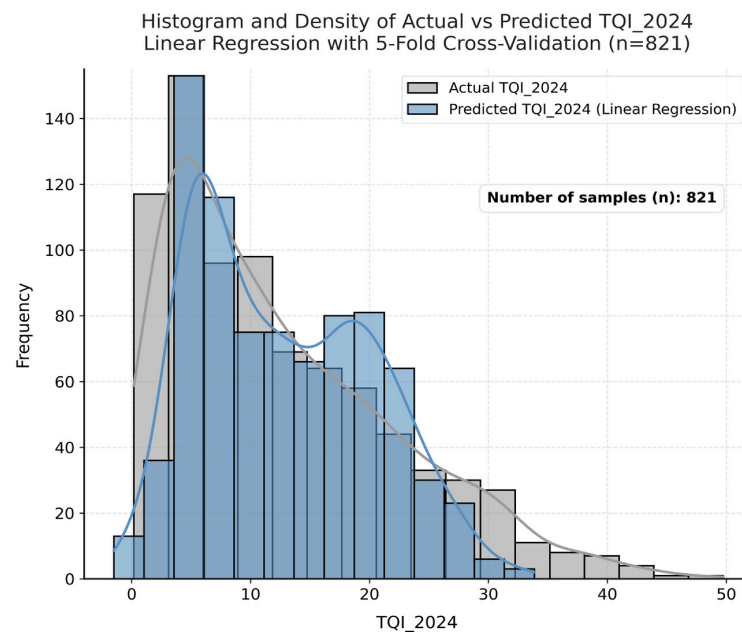


Figure 7. Histogram and density plots of actual vs. predicted TQI_{2024} values obtained with Linear Regression (5-fold cross-validation) where the grey line represents actual data and the blue line represents predicted values. (Source: Author's calculations based on model outputs).

Figure 8 presents a scatter plot of actual versus predicted TQI_{2024} values produced by the RF model. In comparison to Linear Regression, the dispersion around the diagonal is reduced, and a greater proportion of points are located near the ideal fit line ($y = x$) within the TQI range of 5–25. For higher TQI values ($TQI > 30$), isolated outliers are present, but the trend line remains closer to the ideal fit than in the linear model.

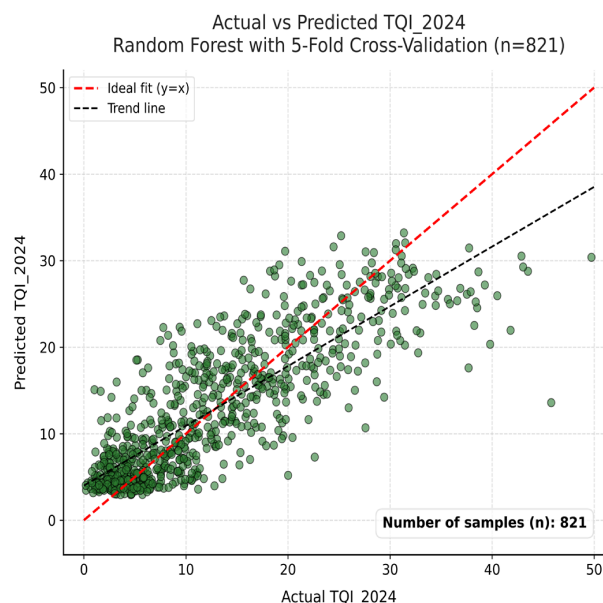


Figure 8. Actual vs. predicted TQI_{2024} values using Random Forest with 5-fold cross-validation, with the red line ($y = x$) indicating the ideal fit. (Source: Author's calculations based on model outputs).

Figure 9 presents a histogram comparison for the RF model. The predicted distribution corresponds closely with the actual distribution across the entire TQI spectrum, including both lower and higher degradation ranges ($TQI > 25$). Minor local deviations persist around the $TQI = 10\text{--}15$ range.

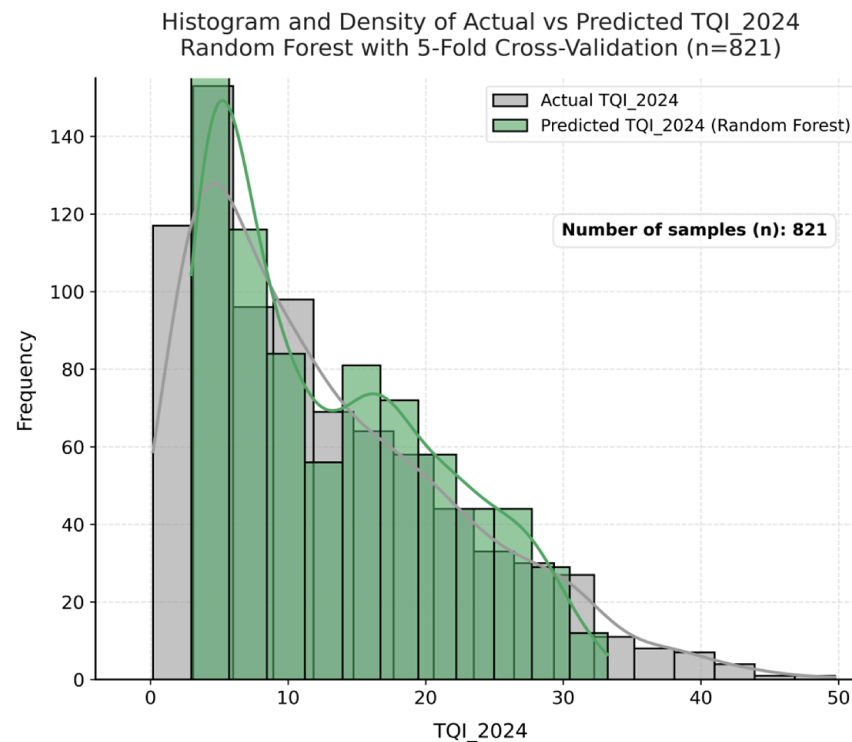


Figure 9. Histogram and density plots of actual vs. predicted TQI_{2024} values obtained with Random Forest (5-fold cross-validation) where the grey line represents actual data and the green line represents predicted values. (Source: Author’s calculations based on model outputs).

4. Discussion

A comparative evaluation of MLR and RF demonstrates distinct theoretical and practical implications for the application of machine learning frameworks in infrastructure asset management.

4.1. Interpretation of Evaluation Metrics

Interpreting statistical evaluation metrics within the context of track quality assessment is essential for understanding the performance of both models.

- **Coefficient of Determination (R^2):** The Linear Regression model produced an R^2 of 0.57, indicating that it explains 57% of the variance in the TQI_{2024} data, while 43% remains unexplained. This outcome demonstrates the model’s limited ability to capture nonlinear processes. In contrast, the RF model achieved an R^2 of 0.69, accounting for nearly 70% of the variance. This improvement underscores the advantages of ensemble methods in analyzing complex systems with nonlinear interactions among geometry, infrastructure components, and operational variables. Furthermore, when implementing a rigorous chainage-blocked cross-validation (GroupKFold grouped by spatial sections) to eliminate spatial autocorrelation along the continuous corridor strictly, the models yielded more conservative performance metrics (R^2 of 0.4111 and RMSE of 7.3703 for Random Forest; R^2 of 0.2658 and RMSE of 8.2299 for Linear Regression). While these spatially blocked scores reflect the inherent challenge of predicting unobserved linear blocks without spatial interpolation bias, they provide a robust,

stress-tested lower bound of model stability, confirming that the ensemble architecture maintains genuine predictive capability across distinct geographical sectors.

- Mean Absolute Error (MAE): The Linear Regression model produced an MAE of 4.68, indicating that predictions deviate by an average of 4.68 units from actual TQI values. Such errors can result in misclassification of priority segments, particularly in critical zones such as tunnels, bridges, and sharp-radius curves. The RF model reduced the MAE to 3.96, representing a 15% improvement and decreasing the likelihood of misclassification, thereby enabling more precise prioritization.
- Root Mean Square Error (RMSE): The Linear Regression model produced an RMSE of 6.30, indicating that significant prediction errors occur frequently. Since RMSE is highly sensitive to extreme deviations, a track segment with an actual TQI of 35 (poor quality) could be predicted as 28.7 (medium quality), leading to an inappropriate downgrade of intervention priority. The Random Forest model reduced the RMSE to 5.35, mitigating the impact of extreme errors and demonstrating greater stability across the entire data range.

4.2. Interpretation and Model Suitability

A comparative evaluation of MLR and RF highlights the advantages of ensemble learning methods in modeling railway track degradation. While MLR provides a transparent and interpretable framework, its strict linearity assumption restricts its capacity to capture complex, nonlinear relationships within the dataset. Railway track degradation is shaped by numerous interacting factors, such as geometric, infrastructural, operational, and maintenance-related variables, which frequently exhibit nonlinear dynamics that linear models are unable to represent adequately.

In contrast, RF offers a distinct advantage in this context. By constructing multiple decision trees and aggregating their outputs, this method effectively captures feature interactions and nonlinear dependencies. This capability results in consistently higher predictive accuracy and closer alignment with the actual distribution of TQI values, as confirmed by statistical metrics and visual analyses. From an engineering perspective, RF provides a more reliable basis for identifying track segments at elevated risk of degradation, thereby supporting proactive maintenance planning.

The algorithmic superiority of the RF model over the MLR baseline is particularly evident when analyzing the compound and overlapping influences present in mountainous railway alignments. In complex corridors such as Kolashin–Podgorica, degradation rarely results from a single parameter but rather from combined spatial stressors. For example, a localized 20-m segment may exhibit both a sharp horizontal curve (radius less than 400 m) and steep longitudinal slopes (greater than 20%) within an unventilated, moisture-retaining tunnel. The parametric structure of MLR is susceptible to variance inflation and cannot disentangle these collinear variables. In contrast, the recursive partitioning mechanism of the RF model effectively captures these compound nonlinear interactions. This spatial adaptability enables the ensemble framework to isolate localized structural transitions, such as rapid ballast stiffness degradation at tunnel portals and bridge approaches, resulting in significantly tighter error bounds in challenging geotechnical environments.

However, MLR retains significant utility. Its simplicity, computational efficiency, and interpretability make it suitable for rapid diagnostic assessments, preliminary evaluations, and scenarios that require transparent relationships between predictors and outcomes. For example, infrastructure managers may use MLR to identify broad trends or conduct sensitivity analyses, while employing RF for more detailed predictive modeling.

In summary, RF offers substantial improvements in predictive performance and generalization, making it a more suitable candidate for integration into predictive maintenance

systems, particularly in geometrically complex contexts such as the Kolashin–Podgorica section. At the same time, MLR remains a practical and complementary approach when transparency, ease of implementation, and rapid interpretation are required.

4.3. Practical Implications for Maintenance Planning

The comparative modeling results have important implications for railway maintenance planning and asset management. The superior performance of the Random Forest model indicates that ensemble-based machine learning approaches yield more reliable forecasts of track geometry quality, particularly in complex environments characterized by frequent tunnels, bridges, and sharp curves. Identifying segments at elevated risk of rapid degradation allows infrastructure managers to prioritize interventions and allocate resources with greater efficiency.

Verification of the operational reliability of the 2024 independent validation profile requires careful examination of the environmental and operational boundary conditions from 2022 to 2024. Field data audits confirm that line speeds, axle-load configurations, and annual cumulative traffic intensity (gross tonnage) remained stable, with no significant operational overhauls or structural disruptions. Therefore, the residual variances observed in the 2024 predictions are attributed to the progressive, natural aging of track components, such as sleeper fatigue and ballast particle abrasion under heavy cyclic loading, rather than to unexpected changes in train operations.

In this context, the feature engineering process aimed to maximize predictive performance and interpretability while ensuring transparency and reproducibility. Domain-relevant features were identified through a review of operational records and track engineering guidelines, with an emphasis on key geometric, infrastructural, operational, and maintenance variables. Candidate features were screened for completeness and relevance, and those with excessive missing data or minimal variance were excluded from the modeling pipeline. Curve radius, steep slopes, component age, and cumulative tamping counts were encoded as continuous variables. Categorical attributes, such as the presence of tunnels, bridges, and stations within each 20-m segment, were encoded as binary indicator variables. To capture temporal degradation effects, dedicated temporal features for each segment were engineered: component age was calculated as the number of years since the last major renewal of key elements (such as sleepers and ballast), and tamping activity history was included as both the cumulative count of tamping operations and the time (in years) since the most recent tamping intervention. All input features were standardized using z-score normalization to aid model training, except for categorical variables, which retained their binary format. Feature selection was further informed by preliminary correlation screening and Random Forest-based feature importance rankings, retaining variables that exhibited meaningful predictive contributions according to both statistical and engineering criteria. Following preprocessing, the finalized feature matrix was integrated alongside static and dynamic attributes, enabling the model to distinguish ongoing degradation trends from random measurement variation and to more accurately reflect the underlying maintenance lifecycle across the evaluation horizon.

Consistent and comprehensive data collection presents a significant challenge in practical deployment. Field measurement campaigns are sometimes delayed, track geometry records may be incomplete, and maintenance records can contain omissions due to manual-entry errors or inconsistencies among operational teams. To address these challenges, infrastructure managers are advised to implement standardized protocols for data collection and reporting, including regular calibration of measurement equipment and digitalization of maintenance logs. The following Data Quality Checklist summarizes

the primary steps required to maintain reliable data inputs for predictive modeling and maintenance planning:

1. Verify calibration certificates for all measurement devices prior to each inspection.
2. Document the time, date, and exact location for every geometry measurement and maintenance activity.
3. Use double-data entry or cross-checks for manual maintenance record logging to reduce potential errors.
4. Apply automated statistical outlier detection routines to raw geometry data after each collection cycle.
5. Manually review flagged anomalies by comparing with the field engineer's notes.
6. Digitize all measurement and maintenance logs, and store them in a centralized and secure database.
7. Implement imputation or use conservative default values for missing or incomplete data segments, based on historical data or engineering judgment.
8. Schedule periodic staff training sessions on data quality assurance procedures.

For engineers seeking to implement a similar predictive framework on other lines, the minimum required data inputs are as follows: (1) periodic track geometry measurements (including TQI or equivalent indices) collected at regular intervals, ideally from track-recording cars; (2) detailed spatial infrastructure parameters such as curve radius, steep slopes, and the location and extent of tunnels, bridges, and stations; (3) operation-related records including line speed, axle-load, and annual gross tonnage for each segment; and (4) maintenance history records, particularly dates and counts of tamping, profiling, and major renewals for sleepers and ballast. For robust model training and validation, at least three inspection campaigns spaced over a multi-year period (4 to 6 years) are recommended to capture both gradual deterioration and the effects of maintenance interventions.

From an operational perspective, improved prediction accuracy reduces the risk of unexpected speed restrictions or safety incidents by enabling earlier detection of critical sections and timely maintenance. This predictive capability decreases reliance on costly emergency interventions and supports service continuity on heavily used corridors. Furthermore, the RF model's capacity to capture nonlinear interactions among geometry, traffic, and maintenance variables supports the development of adaptive maintenance strategies. Multiple linear regression also retains value in practical workflows. Its interpretability enables decision-makers to understand how individual parameters, such as traffic intensity or curvature, influence TQI. This transparency facilitates reporting, regulatory compliance, and communication with non-technical stakeholders. Consequently, multiple linear regression can complement advanced ensemble models by providing diagnostic insights in contexts where simplicity and clarity are required. Utility and clarity are essential.

For infrastructure, it is important for infrastructure managers to recognize that a model with an R^2 of 0.69 leaves a substantial portion of variance unexplained and cannot substitute for safety-critical track-recording car inspections. The model's primary engineering value lies in its role as a network-level predictive screening tool. While the model provides significant insights, several limitations must be acknowledged. Potential sources of error include measurement inaccuracies in track geometry data, inconsistent maintenance records, and temporal gaps in inspection intervals. The model may exhibit bias if certain infrastructure types or operational scenarios are underrepresented in the training data. The specific characteristics of the Kolashin–Podgorica corridor, such as its unique geotechnical and climatic conditions, may not be representative of other railway sections. Dependence on a limited number of inspection campaigns may restrict the model's ability to account for rare events or capture long-term degradation trends. Additionally, the feature set, although comprehensive, may omit relevant contextual variables. For example, variables

such as extreme weather patterns (precipitation, freeze–thaw cycles, temperature extremes), subgrade properties (soil type, drainage conditions, underlying strata), flooding history, and localized environmental hazards are not included in the current model, primarily due to limited data availability or lack of systematic historical records for these factors. The absence of these potentially influential variables may contribute to the observed unexplained variance and limit the detection of certain degradation mechanisms, especially those triggered by environmental or geotechnical events outside the scope of standard track geometry and operational records. Therefore, caution is required when applying the predictive framework to new environments, and local calibration and validation are essential to maintain predictive accuracy. To enhance the model’s practical value and support transferability, a set of concrete steps is recommended for adapting and validating the predictive framework in new railway contexts:

1. Conduct targeted pilot studies by applying the model to additional railway lines that represent a range of geotechnical, operational, and climatic conditions. This involves collecting multi-year monitoring data on track geometry, infrastructure parameters, operational indicators, and maintenance records, in accordance with the methodology established in this study.
2. Collaborate with local infrastructure managers to identify and integrate context-specific variables that may influence degradation patterns, such as local subgrade composition, drainage effectiveness, extreme weather exposure, or unique operational regimes. Including these local factors as additional input features will improve model relevance and predictive accuracy in new environments.
3. Perform local calibration of model parameters and hyperparameters using recent inspection and maintenance data from the target line. This may involve retraining or fine-tuning the model to account for specific degradation mechanisms or data characteristics present in the new context.
4. Implement robust local validation by reserving the most recent available inspection campaign as an independent test set. Compare predicted results with measured values using the same evaluation metrics (R^2 , MAE, RMSE), and assess classification performance at critical intervention thresholds.
5. Review and document any discrepancies between predicted and observed outcomes, and adjust the feature set, preprocessing routines, or model structure as needed to address gaps identified during local validation.
6. After successful validation, incorporate the predictive outputs into local asset management workflows and track maintenance outcomes to enable iterative improvement and feedback-driven adaptation of the model.

By following these steps, infrastructure managers and researchers can ensure that the predictive model is appropriately tailored and validated for each new deployment, thereby supporting robust, data-driven maintenance strategies across diverse railway environments.

For regional railway authorities with limited budgets, this model serves as an automated filter for processing historical geometry logs and rapidly identifying high-risk 20-m clusters along extensive corridors. Early identification and targeted intervention at these hotspots through predictive screening reduces the frequency of urgent, unplanned repairs, which are typically three to five times more expensive than scheduled maintenance. For instance, shifting 10% of emergency repair work to planned maintenance through model-driven screening can yield annual cost savings of 8–12% across maintenance budgets. This resource reallocation not only decreases reactive spending but also optimizes the scheduling and utilization of maintenance crews, supporting more predictable asset management and improved budget justification. Each segment can be incorporated into

maintenance scheduling software, enabling engineers to visualize high-priority clusters on digital track maps. Predictive alerts generated by the model can automatically trigger detailed inspection requests or advanced scheduling of tamping and renewal activities. Additionally, summary reports highlighting emerging hotspots can be generated periodically and distributed to maintenance supervisors, supporting data-driven allocation of crews and resources. This integration facilitates a transition from reactive maintenance based on observed defects to a proactive regime grounded in predictive analytics, thereby improving both cost efficiency and operational reliability.

To further decode the practical asset-management value of the framework, the classification performance shown in Figure 10 must be interpreted through a permanent-way engineering lens. The 2×2 confusion matrix maps the operational intersection between the Random Forest predictive flags and the empirical 2024 ground-truth states across all 821 track segments, categorized by the critical TQI > 30 maintenance intervention threshold.

Operational Confusion Matrix for Track Safety Screening

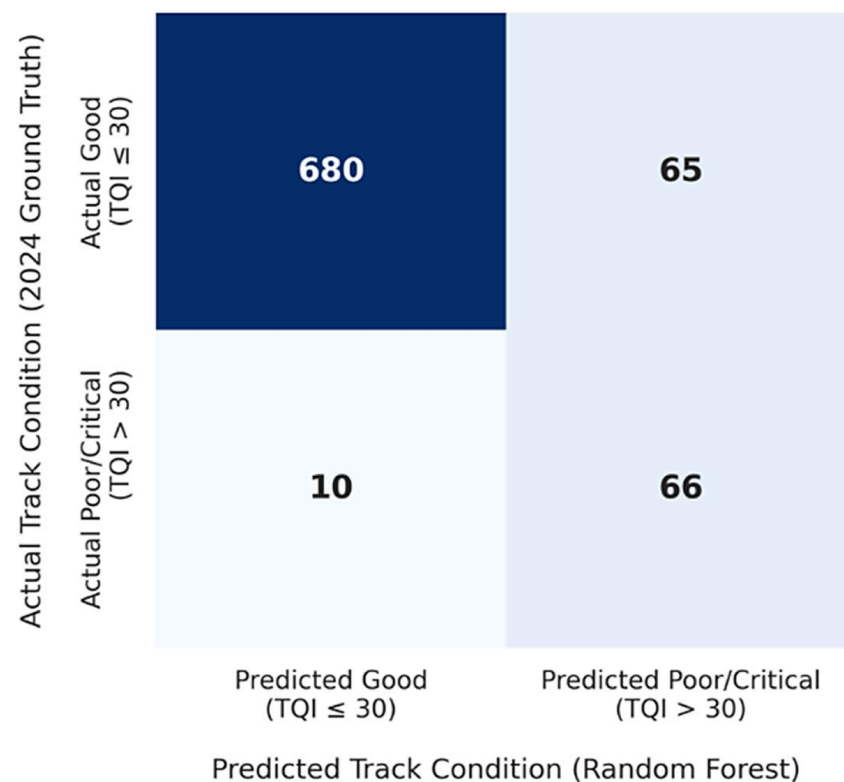


Figure 10. Classification efficiency and predictive screening capability of the Random Forest model were evaluated against the empirical 2024 ground-truth data. (Source: Author’s calculations based on model outputs).

The structural breakdown of the matrix quadrants yields the following engineering implications:

1. True Positives (66 segments): This quadrant represents the core predictive utility for track supervisors. The model successfully identified 66 high-risk 20-m clusters that subsequently crossed into the ‘Poor’ or ‘Critical’ quality zones by 2024. Isolating these structural ‘hotspots’ beforehand allows maintenance engineers to execute targeted local spot-tamping or component tightening, neutralizing geometric acceleration before safety-critical defects occur.
2. True Negatives (680 segments): This represents the vast majority of the corridor footprint in which the track maintained acceptable geometric parameters (TQI ≤ 30).

Correctly mapping these stable zones validates the model's structural filtering accuracy, shielding asset managers from having to deploy heavy machinery to sound-track segments.

3. False Positives/Type I Errors (65 segments): In these locations, the model overpredicted deterioration, flagging them as deficient even though they remained structurally acceptable. From an engineering maintenance standpoint, this constitutes a conservative safety buffer (false alarms); while it schedules redundant field inspections, it introduces no operational hazards to train running.
4. False Negatives/Type II Errors (10 segments): This is the only structurally unfavorable scenario, in which 10 degraded sections breached the $TQI > 30$ limit without triggering a predictive alert. Given the complex geotechnical boundary conditions along mountainous corridors, these localized anomalies can be attributed to sudden subgrade settlement or abrupt ballast fouling that standard cyclic historical trends cannot fully anticipate.

Ultimately, the resulting Sensitivity (Recall) of 86.4% confirms that the ensemble pipeline successfully captures nearly seven-eighths of all impending geometry failures along the network footprint. Paired with a Precision of 82.1% and an exceptional Specificity of 91.3%, the model transforms historical inspection logs into a highly dependable predictive filter, shifting maintenance workflows from reactive crisis management to structured, cost-optimized asset scheduling.

This predictive screening approach optimizes maintenance schedules by reallocating resources from reactive emergency repairs to proactive, planned tamping interventions before geometric defects result in operational restrictions. It provides diagnostic insight and leverages advanced ensemble methods for predictive precision, offering a balanced framework for improving railway maintenance planning. Such an integrated strategy enables railway operators to transition from reactive to proactive asset management, extend track life, optimize maintenance costs, and enhance service reliability.

5. Conclusions

This study systematically compared the predictive performance of Multiple Linear Regression (MLR) and Random Forest (RF) models for forecasting the Track Quality Index (TQI) on the geometrically and geotechnically challenging Kolashin–Podgorica railway section. Analysis of data from three inspection campaigns (2017, 2019, 2022), validated against 2024 measurements, demonstrated that RF consistently surpassed MLR in explanatory power, error reduction, and generalization, as demonstrated by 5-fold cross-validation.

During 5-fold cross-validation, RF consistently achieved an R^2 of 0.69 compared to 0.57 for MLR, along with lower MAE (3.96 versus 4.68) and RMSE (5.35 versus 6.30) values. Furthermore, additional chainage-blocked cross-validation confirmed the underlying spatial stability of the ensemble architecture across unseen geographical blocks. These findings confirm the superior predictive ability of RF in capturing nonlinear degradation patterns.

The analysis highlights several key findings:

- Predictive accuracy: RF achieved higher R^2 values and lower error metrics (MAE and RMSE) than MLR, confirming its greater capacity to capture nonlinear degradation patterns.
- Error distribution: RF more accurately preserved the statistical distribution of TQI values, which is essential for identifying both typical and extreme degradation cases.
- Practical suitability: RF offers a robust and reliable foundation for predictive maintenance planning, particularly in mountainous railway environments characterized by steep slopes, sharp curves, tunnels, and bridges, where degradation dynamics are most pronounced. In contrast, MLR remains valuable as a complementary tool for diagnostic assessments and interpretability.

Integrating ensemble machine learning methods, such as RF, into railway asset management systems enhances proactive maintenance strategies, optimizes resource allocation, and reduces life-cycle costs from an engineering perspective. The scope of this study is limited to a single railway section and a four-period dataset, which constrains the generalizability of the findings. Expanding the dataset to include longer observation periods and additional railway sections across the Western Balkans would further validate the proposed approach. As a practical next step, piloting the predictive model on other lines or regions with differing geological and operational characteristics, such as lowland railways, coastal segments, or urban corridors, is recommended. To ensure feasibility and rigor, future work should employ a structured design specifying: (1) systematic collection of track geometry, infrastructure, operational, and maintenance data aligned with the input requirements described herein; (2) application of standardized data cleaning and segmentation procedures in accordance with established guidelines such as EN 13848 [28] and UIC 518 [29]; (3) adaptation and potential augmentation of model input features to reflect local contextual variables, including subgrade type, climate factors, or regional maintenance practices; and (4) objective evaluation of model performance using predefined validation metrics, such as R^2 , MAE, RMSE, and classification accuracy at critical intervention thresholds. Each pilot study should reserve a portion of data for independent validation to avoid overfitting and ensure unbiased assessment of predictive capacity under real-world conditions. Such pilot studies would demonstrate the model's flexibility and scalability, while also identifying necessary adjustments and calibration requirements for local adaptation. Successful adaptation to other track types or geographical regions may require incorporating additional contextual variables, such as distinct subgrade characteristics, climate patterns, or unique operational regimes. Adjustments to feature selection, training procedures, or the inclusion of supplementary data reflecting variations in track design standards, maintenance practices, and environmental stressors may be necessary to ensure optimal predictive performance. Establishing pilot implementations will facilitate broader adoption and further validate the approach's robustness across diverse railway environments.

Future research should pursue several specific extensions. First, cross-network validation experiments could be conducted by applying the trained model to track sections in different operational and environmental contexts, utilizing both transfer learning and independent retraining to assess model portability. Second, comparative studies involving alternative machine learning algorithms, such as steep slopes boosting, support vector machines, or deep neural networks, would benchmark performance against the RF framework and explore potential improvements in predictive accuracy. Third, controlled field trials or simulation-based experiments, in which maintenance interventions are scheduled based on model forecasts and outcomes are systematically monitored, would provide direct evidence of operational impacts. Additionally, research could focus on developing multi-scale or hierarchical models that combine network-level screening with segment-level diagnosis, and on integrating additional exogenous data sources, including high-frequency sensor readings, climatic data, or real-time traffic information. Addressing these research questions will advance methodological rigor, support continuous model refinement, and align predictive railway maintenance with evolving technological and operational requirements. Although the current model has demonstrated strong predictive power within the specific geotechnical and operational conditions of the Kolashin–Podgorica corridor, its transferability to other contexts, such as lowland plains, arid regions, or heavily trafficked urban rail systems, remains untested. The Random Forest-based framework, with its capacity to capture nonlinear interactions and accommodate a broad range of feature types, is expected to adapt effectively to diverse environments if relevant local variables are incorporated and the model is retrained on representative data. However, variations in track design

standards, environmental stressors, and maintenance practices across different railway networks may influence both the relative importance of features and overall model performance. Therefore, future research should focus on applying and validating the approach across a wider array of geotechnical settings and operational regimes to assess model robustness and identify the need for tailored adjustments in different railway contexts.

Despite the clear benefits of this predictive framework, scaling and deploying it outside the original study area may encounter several practical and organizational barriers. Key challenges include ensuring access to consistent, high-quality data, as many railway networks may lack comprehensive records of track geometry, maintenance history, and infrastructure details. Differences in data collection frequency, measurement standards, and classification schemes could complicate model transferability. Organizational resistance to adopting new data-driven procedures, particularly where maintenance decisions have traditionally relied on expert assessments or legacy systems, may slow implementation. Integration with existing asset management software and workflows often requires technical customization and additional staff training.

Furthermore, securing adequate computational resources and expertise for ongoing model updates, recalibration, and troubleshooting is essential for sustainable deployment. Regulatory approval, data privacy concerns, and the need to demonstrate the model's reliability in different operational contexts may also pose significant hurdles. Addressing these barriers will require not only technical adaptation but also investment in capacity building, clear communication of added value, and strong collaboration between engineering, IT, and management teams.

Although an R^2 of approximately 0.69 indicates strong alignment with global trends, residual unexplained variance remains. This uncertainty is primarily due to highly localized, stochastic track defects, such as sudden ballast fouling or isolated mechanical fastener fatigue, which occur at scales smaller than the 20-m evaluation window. Addressing this variance in future studies may require higher-resolution monitoring data, such as more granular track-inspection intervals or the integration of advanced sensor technologies, including distributed acoustic sensing or continuous axle-box acceleration monitoring. Incorporating external environmental factors, such as localized precipitation data or subgrade moisture measurements, may further enhance model fidelity in capturing unpredictable triggers of degradation. Developing hybrid modeling frameworks that combine the interpretability of statistical approaches with the adaptive learning capacity of advanced machine learning architectures, such as deep neural networks or steep slopes boosting methods, represents a promising direction for reducing residual error and better capturing rare or complex defect patterns. The proposed framework should be regarded as a high-fidelity, network-level predictive screening tool rather than a safety-critical substitute for high-frequency manual track-recording inspections. Its primary engineering value lies in isolating high-risk corridors, thereby enabling a shift from reactive crisis management to proactive, cost-optimized asset scheduling.

Future work should explore hybrid modeling approaches that combine the interpretability of statistical methods with the predictive power of advanced machine learning and extend the analysis to broader railway networks.

In conclusion, ensemble-based predictive modeling offers a robust and scalable foundation for data-driven railway infrastructure management, supporting safer, more resilient, and economically sustainable railway networks.

Author Contributions: Conceptualization, I.N.T. and Z.Z.; methodology, I.N.T.; software, I.N.T.; validation, I.N.T., Z.Z. and J.D.; formal analysis, I.N.T. and R.R.; investigation, I.N.T. and V.G.; resources, I.N.T., R.R. and V.G.; data curation, I.N.T.; writing—original draft preparation, I.N.T.; writing—review and editing, I.N.T., Z.Z. and J.D.; visualization, I.N.T. and R.R.; supervision, Z.Z.;

project administration, I.N.T. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to institutional restrictions, but are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the faculty and technical staff for providing support and administrative assistance during this study. During the preparation of this manuscript, the authors used Grammarly Pro (web version 14.1314.0) for the purposes of text structuring, styling consistency, and language editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Al-Douri, Y.; Tretten, P.; Karim, R. Improvement of railway performance: A study of Swedish railway infrastructure. *J. Mod. Transp.* **2016**, *24*, 22–37. [\[CrossRef\]](#)
2. Paixão, A.; Fortunato, E.; Calçada, R. Smartphone's sensing capabilities for on-board railway track monitoring: Structural performance and geometrical degradation assessment. *Adv. Civ. Eng.* **2019**, *2019*, 1729153. [\[CrossRef\]](#)
3. Kampczyk, A. Measurement of the geometric center of a turnout for the safety of railway infrastructure using MMS and total station. *Sensors* **2020**, *20*, 4467. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Khosravi, M.; Ahmadi, A.; Kasraei, A.; Nissen, A. Optimisation of railway tamping scheduling. *Heliyon* **2024**, *10*, e40844. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Charoenwong, C.; Connolly, D.; Woodward, P.; Galvín, P.; Costa, P. Modelling railway track differential settlement for prediction of future deterioration and maintenance intervals. *Comput. Concr. Compos.* **2020**, *1*, 9. [\[CrossRef\]](#)
6. Shang, Y.; Nogal, M.; Wang, H.; Wolfert, A. Systems thinking approach for improving maintenance management of discrete rail assets: A review and future perspectives. *Struct. Infrastruct. Eng.* **2021**, *19*, 197–215. [\[CrossRef\]](#)
7. Litherland, J.; Andrews, J. A Petri net methodology for modelling the maintenance of railway route sections. In Proceedings of the 29th European Safety and Reliability Conference (ESREL), Hannover, Germany, 22–26 September 2019; pp. 452–459. [\[CrossRef\]](#)
8. Xu, P.; Liu, R.; Wang, F.; Wang, F.; Sun, Q. Railroad track deterioration characteristics based on track measurement data mining. *Math. Probl. Eng.* **2013**, *2013*, 970573. [\[CrossRef\]](#)
9. Widayastuti, H. Prediction of speed limit on the railway track using track quality index and multibody dynamics simulation. In Proceedings of the 2023 International Conference on Railway Engineering, Beijing, China, 12–15 October 2023; Volume 33, pp. 45–52. [\[CrossRef\]](#)
10. Krakutovski, Z. *Study of Railway Track Renewal of Corridor X in Republic of Macedonia: Technical Preparation and Economic Assessment*; Faculty of Civil Engineering: Skopje, North Macedonia, 2009.
11. South East Europe Transport Observatory (SEETO). *Report on Rail Maintenance on TEN-T Core Network in Western Balkans*; SEETO: Belgrade, Serbia, 2016.
12. Jovanović, S.; Guler, H.; Čoko, B. Analiza degradacije kolosijeka u sklopu sustava za upravljanje održavanjem željezničke infrastrukture. *Gradevinar* **2015**, *67*, 247–258. [\[CrossRef\]](#)
13. Khajehei, H.; Ahmadi, A.; Soleimanmeigouni, I.; Haddadzade, M.; Nissen, A.; Latifi Jebelli, M.J. Prediction of track geometry degradation using artificial neural network: A case study. *Int. J. Rail Transp.* **2022**, *10*, 24–43. [\[CrossRef\]](#)
14. Wardahni, N.; Latief, Y. Railway infrastructures e-maintenance systems: A literature review. *IOP Conf. Ser. Earth Environ. Sci.* **2024**, *1324*, 012045. [\[CrossRef\]](#)
15. Yokoyama, A. Innovative changes for maintenance of railway by using ICT—To achieve ‘smart maintenance’. *Procedia CIRP* **2015**, *38*, 24–29. [\[CrossRef\]](#)
16. Goyal, P.; Chan, A.; Jaiswal, N. Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmos. Environ.* **2006**, *40*, 2068–2077. [\[CrossRef\]](#)
17. McFarlane, C.; Raheja, G.; Malings, C.; Appoh, E.; Hughes, A.; Westervelt, D. Application of Gaussian mixture regression for the correction of low-cost PM2.5 monitoring data in Accra, Ghana. *ACS Earth Space Chem.* **2021**, *5*, 2268–2279. [\[CrossRef\]](#)
18. Setiyorini, T.; Frieyadie, F. Comparison of linear regressions and neural networks for forecasting electricity consumption. *J. Pilar Nusa Mandiri* **2020**, *16*, 135–140. [\[CrossRef\]](#)

19. Papalexopoulos, A.; Hesterberg, T. A regression-based approach to short-term system load forecasting. In Proceedings of the Power Industry Computer Application Conference, Seattle, WA, USA, 1–5 May 1989; pp. 414–423. [\[CrossRef\]](#)
20. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [\[CrossRef\]](#)
21. Bagui, S.; Bennett, T. Optimizing random forests: Spark implementations of random genetic forests. *Balk. J. Electr. Comput. Eng.* **2022**, *1*, 44–52. [\[CrossRef\]](#)
22. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
23. Burzanović, S.; Kovačević, B.; Marović, B.; Miljić, M.; Rastoder, Š. *100 Godina Željeznice Crne Gore*; Željeznica Crne Gore: Podgorica, Montenegro, 2009.
24. Wang, H.; Markine, V. Modelling of the long-term behaviour of transition zones: Prediction of track settlement. *Eng. Struct.* **2018**, *156*, 294–304. [\[CrossRef\]](#)
25. Ehrhart, U.; Knabl, D.; Marschnig, S. Track deterioration model—State of the art and research potentials. *Infrastructures* **2024**, *9*, 86. [\[CrossRef\]](#)
26. Li, W.; Pu, H.; Schonfeld, P.; Song, Z.; Zhang, H.; Wang, L.; Peng, L. A method for automatically recreating the horizontal alignment geometry of existing railways. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *34*, 71–94. [\[CrossRef\]](#)
27. Offenbacher, S.; Koczwar, C.; Landgraf, M.; Marschnig, S. A methodology linking tamping processes and railway track behaviour. *Appl. Sci.* **2023**, *13*, 2137. [\[CrossRef\]](#)
28. EN 13848-1:2003+A1; Railway Applications—Track—Track Geometry Quality—Part 1: Characterization of Track Geometry. CEN: Brussels, Belgium, 2008.
29. International Union of Railways (UIC). *UIC Code 518: Testing and Approval of Railway Vehicles from the Point of View of Their Dynamic Behaviour—Safety—Track Fatigue—Ride Quality*, 2nd ed.; UIC: Paris, France, 2009. Available online: <https://uic.org> (accessed on 15 May 2026).
30. Fu, G.; Wilkinson, S.; Dawson, R.J. A spatial network model for civil infrastructure system development. *Comput.-Aided Civ. Infrastruct. Eng.* **2016**, *31*, 491–510. [\[CrossRef\]](#)
31. Andrienko, N.; Andrienko, G.; Höchtl, R.; Wrobel, S. Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics. *Inf. Syst.* **2016**, *59*, 142–157. [\[CrossRef\]](#)
32. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [\[CrossRef\]](#)
33. Farrar, C.R.; Worden, K. *Structural Health Monitoring: A Machine Learning Perspective*; Wiley: Chichester, UK, 2012. [\[CrossRef\]](#)
34. Bishop, B.C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006. Available online: <https://link.springer.com/book/9780387310732> (accessed on 15 May 2026).
35. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009. [\[CrossRef\]](#)
36. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 15. [\[CrossRef\]](#)
37. EN 13848-6:2026; Railway Applications—Track—Track Geometry Quality—Part 6: Determination of Track Geometry Quality Levels. European Committee for Standardization: Brussels, Belgium, 2026. Available online: <https://standards.iteh.ai/catalog/standards/cen/accf190b-6761-4e5e-9146-dbe1a2b47e53/en-13848-6-2026> (accessed on 15 May 2026).
38. Vale, C.; Ribeiro, I.M. Railway condition-based maintenance model with stochastic deterioration. *J. Civ. Eng. Manag.* **2014**, *20*, 686–692. [\[CrossRef\]](#)
39. Yan, T.; Corman, F. Assessing and Extending Track Quality Index for Novel Measurement Techniques in Railway Systems. *Transp. Res. Rec.* **2020**, *2674*, 148–159. [\[CrossRef\]](#)
40. CEN/TC 256/WG 1; Railway Applications—Track—Track Geometry—Part 1: Geometric Design of Track. CEN: Brussels, Belgium, 2008.
41. Soleimanmeigouni, I.; Ahmadi, A.; Khouy, I.A.; Letot, C. Evaluation of the effect of tamping on the track geometry condition: A case study. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *232*, 408–420. [\[CrossRef\]](#)
42. Aseer, S.; Onsy, A.; Varley, R.H. Condition monitoring of railway infrastructure: A case study of a light rail system. *arXiv* **2024**. [\[CrossRef\]](#)
43. Andrén, A.; Dahlström, L.; Nordlund, E. Field observations of water and ice problems in railway tunnels from a maintenance perspective. *J. Earth Sci. Geotech. Eng.* **2023**, *13*, 11–58. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Iordan, P.; Mihaela, S.O.; Marian, S.R. Dynamic behavior of railway bridges under high-speed traffic loads. *Rom. J. Transp. Infrastruct.* **2024**, *13*, 1–16. [\[CrossRef\]](#)
45. Ivina, D.; Ma, Z. Stability assessment of railway trackwork scheduling in Sweden. *Eur. Transp. Res. Rev.* **2024**, *16*, 12. [\[CrossRef\]](#)
46. Mirković, N.; Mirković, U.; Stevanović, D. Requirements for train–track–bridge dynamic interaction analysis. *Proceedings* **2025**, *1*, 141–151. [\[CrossRef\]](#)

47. Wang, J.; Chen, X.; Li, X.; Wu, Y. Influence of heavy haul railway curve parameters on rail wear. *Eng. Fail. Anal.* **2015**, *57*, 511–520. [\[CrossRef\]](#)
48. Makarov, Y.; Zaleskyi, R.; Mykhalichenko, M. Influence of different factors on the value of the rail wear rate. *Acta Tech. Jaurinensis* **2024**, *17*, 45–58. [\[CrossRef\]](#)
49. Zakeri, J.A.; Yousefian, K. Experimental investigation into the longitudinal resistance of ballasted railway track. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2020**, *235*, 969–981. [\[CrossRef\]](#)
50. Li, X.; Ekh, M.; Nielsen, J.C.O. Three-dimensional modelling of differential railway track settlement using a cycle domain constitutive model. *Int. J. Numer. Anal. Methods Geomech.* **2016**, *40*, 1758–1770. [\[CrossRef\]](#)
51. Fattah, M.Y.; Majeed, Q.G.; Joni, H.H. Effect of degree of saturation on stresses and pore water pressure in the subgrade layer caused by railway track loading. *Railw. Sci.* **2024**, *3*, 413–436. [\[CrossRef\]](#)
52. Costa, J.N.; Antunes, P.; Magalhães, H.; Pombo, J.; Ambrósio, J. A novel methodology to automatically include general track flexibility in railway vehicle dynamic analyses. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2020**, *235*, 478–493. [\[CrossRef\]](#)
53. Indraratna, B.; Nimbalkar, S.; Anantanasakul, P.; Rujikiatkamjorn, C.; Neville, T. Performance monitoring of rail tracks stabilized by geosynthetics and shock mats: Case studies at Bulli and Singleton in Australia. In Proceedings of the Geo-Congress 2013, San Diego, CA, USA, 3–6 March 2013; pp. 19–33. [\[CrossRef\]](#)
54. Elkhoury, N.; Hitihamillage, L.; Moridpour, S.; Robert, D. Degradation prediction of rail tracks: A review of the existing literature. *Open Transp. J.* **2018**, *12*, 88–104. [\[CrossRef\]](#)
55. Majstorović, I.; Ahac, M.; Madejski, J.; Lakušić, S. Influence of the analytical segment length on the tram track quality assessment. *Appl. Sci.* **2022**, *12*, 10036. [\[CrossRef\]](#)
56. Andrews, J.; Prescott, D.; Rozières, F.D. A stochastic model for railway track asset management. *Reliab. Eng. Syst. Saf.* **2014**, *130*, 76–84. [\[CrossRef\]](#)
57. Prescott, D.; Andrews, J. Investigating railway track asset management using a Markov analysis. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2013**, *229*, 402–416. [\[CrossRef\]](#)
58. Zhou, T.; Hu, B.; Xu, P.; Yan, B. Numerical study on the evolution of force chain inside railway ballast under tamping operation. *Appl. Mech. Mater.* **2015**, *724*, 275–278. [\[CrossRef\]](#)
59. Barbir, O.; Adam, D.; Kopf, F.; Pistol, J.; Antony, B.; Auer, F. In-situ ballast condition assessment by tamping machine integrated measurement system. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *710*, 012071. [\[CrossRef\]](#)
60. Abbasi, A.; Zakeri, J.A.; Norouzi, E.; Mosayebi, S.A. Field investigation on the effect of the tamping machine and dynamic track stabilizer on changing the rail support modulus. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2024**, *238*, 1072–1083. [\[CrossRef\]](#)
61. Faris, M.; Núñez, A.; Su, Z.; De Schutter, B. Distributed optimization for railway track maintenance operations planning. In Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 1240–1246. [\[CrossRef\]](#)
62. Rosati, S.; Balestra, G.; Knaflitz, M. Comparison of Different Sets of Features for Human Activity Recognition by Wearable Sensors. *Sensors* **2018**, *18*, 4189. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Aksoy, S.; Haralick, R.M. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognit. Lett.* **2001**, *22*, 563–582. [\[CrossRef\]](#)
64. Al-Mahrami, N.; Jabri, N.A.; Sallam, A.A.W.; Jahdhami, N.A.; Zadjali, F. ChromoCheck: Predicting Postnatal Chromosomal Trisomy Cases Using a Support Vector Machine Learning Model. *Genes* **2025**, *16*, 695. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Cai, Z.; Zafferani, M.; Akande, O.M.; Hargrove, A.E. Quantitative structure–activity relationship (QSAR) study predicts small-molecule binding to RNA structure. *J. Med. Chem.* **2022**, *65*, 7262–7277. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Jin, S.; Zhang, Y.; Chen, Y.; Wang, Y. A smoothness optimization method for horizontal alignment considering ballasted track maintenance. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *38*, 739–761. [\[CrossRef\]](#)
67. Lu, Y.; Zhang, J.; Chen, J. Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost UAV system. *Plant Methods* **2019**, *15*, 45. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Resti, Y.; Irsan, C.; Latif, J.F.; Yani, I.; Dewi, N.R. A Bootstrap-Aggregating in Random Forest Model for Classification of Corn Plant Diseases and Pests. *Sci. Technol. Indones.* **2023**, *8*, 288–297. [\[CrossRef\]](#)
69. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [\[CrossRef\]](#)
70. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, É. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830. [\[CrossRef\]](#)
71. Chatelain, Y.; Yong, N.; Kiar, G.; Glatard, T. PyTracer: Automatically profiling numerical instabilities in Python. *arXiv* **2021**, arXiv:2112.11508. [\[CrossRef\]](#)
72. McKinney, W. Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 56–61. [\[CrossRef\]](#)

73. Miles, J.N.V. R squared, adjusted R squared. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2014; Volume 1, pp. 1–5. [[CrossRef](#)]
74. Renaud, O.; Victoria-Feser, M. A robust coefficient of determination for regression. *J. Stat. Plan. Inference* **2010**, *140*, 1852–1862. [[CrossRef](#)]
75. Thorn, P.D. A formal solution to Reichenbach’s reference class problem. *Dialectica* **2019**, *73*, 349–366. [[CrossRef](#)]
76. Prins, S.C.B.; Smith, E.P. Using biological metrics to score and evaluate sites: A nearest-neighbour reference condition approach. *Freshw. Biol.* **2006**, *52*, 98–111. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.