An artificial intelligence approach for automated asset management of railway systems

Luca Di Costanzo
Department of Civil, Architectural and Environmental Engineering
University of Naples "Federico II"
Naples, Italy
luca.dicostanzo@unina.it

Angelo Coppola
Department of Civil, Architectural and Environmental Engineering
University of Naples "Federico II"
Naples, Italy
angelo.coppola@unina.it

Stefano Marrone
Department of Electrical and ICT Engineering
University of Naples "Federico II"
Naples, Italy
stefano.marrone@unina.it

2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI) | 979-8-3503-6213-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/RTSI61910.2024.10761591

***Abstract*—Automated diagnostic and predictive asset management capabilities are of paramount importance in the era of connected and automated cooperative mobility. A diagnostic vehicle can scan the rail network and process sensor measurements to prevent incoming disruptions and ensure smooth operation of automated transportation services. This requires the development of reliable algorithms that enable early warning and predictive asset management. An algorithm based on artificial intelligence techniques is presented here. The algorithm analyses diagnostic measures and relates them to observed faults on the rail network. In operation mode, the algorithm predicts maintenance needs based on current measurements.**

***Index Terms*—Asset Management, Automated Diagnosis, Diagnostic Train, railway, RUL, Railway, ML, Machine Learning**

I. Introduction

Dynamic scheduling of maintenance with a predictive approach is a key task of an asset management process oriented to optimize the usability of the infrastructures and services. In case of railway networks, large-scale tasks, such as grinding, tamping, and other track geometry maintenance, have special requirements in terms of cost, (un)availability of track, disruption of services, loss of quality of service, and other challenges for scheduling and business management [1]. Maintenance tasks to be carried out frequently or requiring a long maintenance time window have the most significant effect on track availability and network capacity. To optimize asset management, algorithms have been proposed aimed at maximizing track availability and, in general, minimizing the generalized cost of maintenance [2]. As a general consideration from relevant research, it is worth saying that preventive maintenance problems in large-scale rail networks involve hundreds of assets and very complex relationships between them, which generates a very large number of constraints [3]. In all cases, an optimization strategy can only be based on track defect detection and prediction, with a paradigm shift from corrective or conditional maintenance toward predictive maintenance. This allows for maintenance planning and for application of bi-level programming for optimal re-routing of services and advanced traveller information, as it happens in road contexts [4]. It is therefore crucial to develop and use a degradation prediction model and/or a remaining life anticipation model that, with reference to track geometry measurements and fed by diagnostic train data, is able to anticipate the need for intervention at least 60-90 days in advance, in order to have enough time to use optimization algorithm and put into action the best solution.

The work has been partially carried out within the MOST – Sustainable Mobility National Research Center and received funding from the European Union Next-GenerationEU (PNRR, Missione 4.2.1.4 – D.D. 1033 17/06/2022, CN00000023). This manuscript reflects only the authors' views and opinions. Neither the European Union nor the European Commission can be considered responsible for them.
Moreover, it has been partially funded by the Ministry of University and Research (Ministero dell'Università e della Ricerca) under the project DIGIT-CCAM (grant E67G22000010005).

II. Methodology

One of the most relevant aspects of rail asset management is the maintenance of optimal track conditions. Indeed, the materials of the railway track, and thus the track geometry, degrade due to repeated loads from passing trains and environmental conditions. On the other hand, the geometry must meet specific and very stringent quality requirements, both to avoid speed limitations and to reduce the risk of derailment. It is therefore inspected periodically to detect defects before they reach unacceptable operating levels. A very critical issue affecting the stability and longevity of rail infrastructure is the degradation of ballast, consisting of crushed stone or gravel. This latter break down and lose its structural integrity, resulting in smaller particles, or fines, which can lead to safety and efficiency problems [5], [6]. For instance, repeated loads cause vertical and lateral deformation, reduced drainage and track misalignment [7]. Moreover, the accumulation of fines can also create a more rigid structure, diminishing the ballast's ability to distribute loads and absorb impacts. Regular maintenance, including cleaning or replacing the ballast, is essential to mitigate these effects and ensure safe and efficient railway operations Thus, degradation of track geometry is influenced by several factors, among others: traffic loads and speeds, materials and construction methods, maintenance history. In order to continuously measures the track geometry, several parameters are observed, among others: vertical alignment (or longitudinal level), horizontal alignment, gauge, and skew. The standards prescribe minimum and maximum allowable values for these parameters, depending on the type of railway line. If beyond these limits, various actions must be taken, which can range from immediate action, to corrective maintenance, to warning that requires attention to the situation with a view to planning for non-immediate action.

A. Approach and hypotheses

In this paper, the feasibility of using diagnostic train data to identifying the useful remaining time before the need (from standards, regulations, and guidelines) to take action. We refer, then, to so-called Remaining Useful Life (RUL) techniques. The analyses of this study are based on measured track geometry data obtained from diagnostic train passes. Given the nature of such survey and measurements, the dataset used is also referred to as mobile data. The hypotheses under which the system is modelled are as enlisted below:

- The deterioration process is designed for discrete time intervals. Even though the deterioration of track geometry is continuous over time, it is assumed that it is possible to sample the condition of the track with an inter-time Δt dependent on the frequency of passing of diagnostic trains.
- The state of degradation is represented by the dynamics of the measurements themselves; a system in which one or more of the measurements have gone outside the predetermined limits of tolerance is considered degraded.
- The degradation process can be interpreted from one sampling interval to the next; the state of the system depends on explanatory variables, among which the most important is assumed to be the cumulative traffic load on the railroad track.

The work has been carried out with reference to data provided by one of the main railway operator in Italy; the access to data is not direct but provided off-line. The data sources used were heterogeneous and included: i) the data lake of diagnostic measurements (with reference to track geometry); ii) the database of work orders and condition maintenance interventions; iii) rail traffic data; iv) weather and environmental data (forecasts). The work refers to a feasibility level, that is, with the aim of analysing the quality of the data in terms of their possible use for the realization of the considered RUL-oriented predictors.

B. Reference Methodologies

In recent years, predictive technologies have increasingly relied on techniques based on big data analysis to monitor, possibly in real time, the health of a system in order to improve the accuracy and reliability in predicting failures and/or malfunctions [8]. The use of these technologies responds to the need to analyse and process information from multiple and heterogeneous data sources (temperature, vibration, sensor measurements, etc.) in order to consider a vast number of possible causes and concomitant causes of failure so that maintenance strategies deemed most appropriate can be implemented. The term Artificial Intelligence (AI) encompasses the set of techniques, applications, and fields of research related to the study of systems (hardware, software, or hybrids) that accomplish tasks that would require the use of intelligence if performed by humans [9], [10]. Among its subdomains, Machine Learning (ML) [11] is the one most applied in the field of industrial predictive maintenance, due to its characteristic of learning by examples the task to be performed. In the RUL field, Machine Learning is aimed at tuning Regression Models, designed to predict the remaining life of a piece of equipment. An example of this scenario is the numerical prediction of the days to failure of a piece of machinery, from the analysis of measurements (current and historical) taken on the machinery itself. It is worth noting that the use of ML techniques for predictive maintenance is increasingly present due to the amount of data produced daily by control systems [12], [13]. Indeed, the presence of data acquired both during nominal operation of a system and in the presence of abnormal operation, lays the foundation for learning mathematical models designed to estimate the occurrence of failures, based on the analysis of the data itself [14]. Specifically, once the necessary data has been obtained and properly cleaned, a ML model can be trained by going on to identify which variables to include, through a process known as feature engineering. Not carrying out this phrase of skimming is a common mistake, based on the idea that having more features to feed to an algorithm yields better results. Instead, it has been repeatedly shown that in the presence of a non-unlimited number of data (i.e., in all real-world cases) as the number of features involved increases, the predictive ability of the models tends to deteriorate, due to the so-called curse of dimensionality.

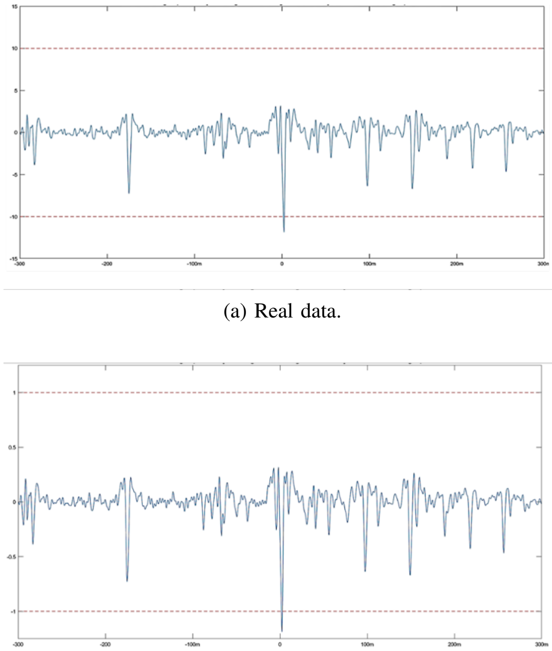
III. PREPARATION OF DATA AND PRELIMINARY ANALYSES

The activities on mobile data (gathered with diagnostic trains) include an important initial phase of data extraction from the railway operator database. We have adopted as a starting point for extraction the identification id from the database of the Work Orders (WOs in the following), that is the maintenance works executed on tracks. They represent both the outcome of a system failure as measured by the indicator we would like to predict, and a change in the state of track geometry (typically restored to new, or otherwise subjected to partial restoration with reference to known geometric characteristics). About 290 intervention points were identified from WOs. For each of the intervention points, all the measures of the diagnostic train passages backward over time (up to a possible previous WO) have been considered for a range of track of 300 meters around the intervention point. This has allowed the identification of track segments of diagnostic measures included between two interventions on the same section of the railway, i.e. between two refurbishment of the section and therefore discontinuities in the degradation process and instants of RUL extension. Each time interval obtained can therefore be considered, barring interventions not recorded, not surveyed, or contained in DBs other than those interrogated, an uninterrupted series of measurements not affected by partial or total geometry renewal. The series were in number of 464, that means that on average 60% of the track segments were effected by more than one renewal and for them a complete set of measures is available between two degradation points. For each series of measures, also non-geometric and non-maintenance data have been retrieved from the database of the railway operator, such as the train traffic (passages of non-diagnostic trains) and weather conditions. It is worth noting that 392 of the 464 series (85%) were terminated by a triggering condition that was an inadmissible value for the vertical alignment (longitudinal level), beyond the threshold value. This must be taken into account when the capabilities of different predictors is calibrated, as in practice the longitudinal level predictor has the power to predict 85% of the observed ends of RUL. The resulting WOs were almost all consisting in ballast intervention (with tamping machines) and the likely phenomenon at the basis of the predictor activation is deterioration, mostly due to densification, distortion, and degradation phenomena of the railway superstructure.

A. Selection and normalization of the measures

All the data gathered by accessing the database of the railway operator were referred to a line with a travel speed lower than 250 km/h. In view of this occurrence and on the expert knowledge of the authors, the analysis was restricted to only four of the measures taken by the diagnostic trains. The considered candidate predictors are: Gap; Longitudinal level; Transverse level; Skew. The normalization process is applied in the range of 300 meters before and after each intervention point (600 meters in total). It consists in simply rescaling the range of the measures with reference to the range of acceptable values, that is, the one that do not trigger the maintenance warning. For instance, assume the considered measure is the longitudinal level and that it can be tolerated if in the range $\pm 10mm$ around the nominal design value (that is zero). Then, the positive measured longitudinal values are scaled by 10 and the negative ones by 10 as well. An example of real data for the longitudinal level before and after the normalization is given in Figure 1a and Figure 1b below. Following the normalization phase of the measurements, a set of features, or rather statistical variables, were extracted to express, with a single value, the performance of the attribute in the entire geometric surroundings of the intervention point. In this way, a time series of spatial profiles is transformed, for each measure, in a time series of numerical values for the candidate predictor. The considered features are: Mean Value; Median value; Kurtosis.

All measures above are popular and have a well-known meaning, but the kurtosis, which is less frequently adopted. Easily speaking, the kurtosis can be considered as the measure of the discrepancy of a statistical distribution with respect to a normal distribution. If zero, such a discrepancy is null, if less than zero the distribution is more uniform than the Normal one (with the same variance), if greater than zero it is more concentrated than the corresponding Normal distribution. In our case, a high value for the kurtosis explain the presence in the observed railway section around the intervention point of particularly higher (than the mean) values of the measure in some points in the considered section. The evaluation of all features listed above has been also calculated for a restricted surrounding of the intervention point, measuring 200 meters instead of 600 meters.



(a) Real data.

(b) Normalised data.

Fig. 1: Longitudinal level around an intervention point, considering a range of $\pm$ 300 meters. The horizontal dotted red lines indicate acceptance range of $\pm$ 10 mm around the nominal design value 0.

IV. MODEL SELECTION, TRAINING AND RESULTS

In our study case, the aim is to obtain a ML model that, based on the features elaborated on the measures taken by a diagnostic train over time, is able to predict the number of days remaining for the end of useful life (RUL) of the measured section of railway. The model will be trained with respect to observed measures and known (observed) remaining days at the moment the measures are taken. As any ML model, it will be trained against observed data and will be evaluated against a different set of observed data of the same kind. The measure(s) able to conciliate predicted remaining days with observed ones are called predictor(s). Finally, the predictor(s) and the ML model are able to forecast, given a set of measures repeated over time, the RUL, that is the number of days left before an intervention is required because of the failing of the measure(s) in the non-acceptable region of values.

A. The adopted model

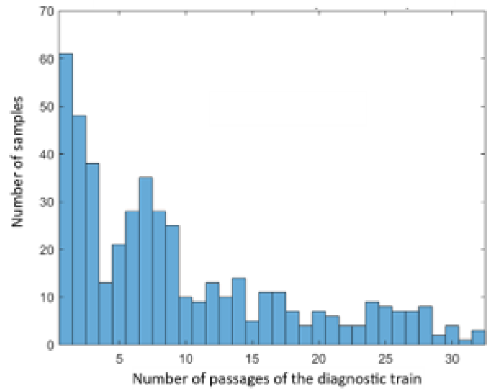
The nature of the use case and of the available data samples suggest adopting a Convolutional Neural Networks (CNN) [15], a family of neural networks that can autonomously learn how to extract information from the data. In this case, the use of the convolution operator along the temporal component allows learning spatio-temporal information, which is useful for predictive purposes. After some preliminary tests aimed at designing the structure of the network, the model consists of a CNN having two convolutional layers (each with 10 filters of size 3×1, stride 1 and no padding) and two dense layers (each having 100 neurons), interspersed with a dropout layer (set with a probability of 0.7). The activation function used was the ReLu, the batch size was set to 32 and the learning rate to 0.01, with a linear decay of 0.001. The maximum number of epochs was set to 25. At a given time an input sample for training the model is a matrix with as many columns as the considered measures (Gap, Longitudinal Level, Transvers Level, Skew), elaborated toward the chosen feature (Mean or Median, or SDV or Kurtosis) and as many rows as the number of passages over time available for the railway section at the considered time. The output is the number of remaining days (RUL).

B. Model training and results

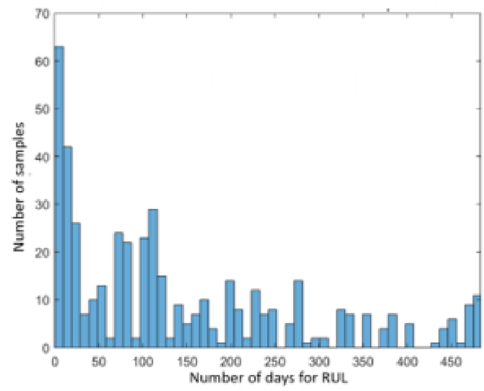
The model has been trained with respect to the data measured by diagnostic train. In order to isolate conditions when the tracks start from their full functionality and progressively degrade toward the maintenance intervention, the measures from a major intervention to another have been considered. The outcome is that for different intervention point the timespan of the measures is different, as it is different the time elapsed from one major maintenance to another.

As the elapsed time is different, the number of diagnostic train passages in that time is different, ranging from few (more fragile track points) to several. Figure 2a shows the distribution of the number of diagnostic train passages among the considered intervention points. As the passages of the diagnostic train are not constant over the years and the months, the distribution of the number of passages has a different but consistent shape in terms of number of RUL days observable at each passage of the diagnostic train, which is depicted in Figure 2b. An example of results in terms of observed and forecasted (by the ML model) RUL days is depicted in Figure 3.

The overall results in terms of prediction ability of the trained model, for all the sample adopted for the validation of the model, are presented in Table I below by using standard error indicators as MAE and RMSE. The spatial range (600 m vs 200 m) has a limited impact on prediction based on AI techniques, but has a high impact on the stage of choosing the most effective features. Therefore, it is suggested to explore (i) larger spatial ranges and (ii) additional aggregation techniques (e.g., symmetry). To analyse the impact of the number of temporal acquisitions (i.e., diagnostic train passes) on the predictive ability of the model, we report in Figure 4 the error trend according to the number of available temporal samples. The figure show a clear decreasing trend in the error, which tends to decrease as the number of available time instants increases. Note in particular that, for samples having a number of temporal acquisitions less than 10, the model tends to underestimate the RUL (i.e., it indicates the need for maintenance earlier than indicated by the measured RUL), while it tends to more closely follow the measured RUL for samples ahead a number of temporal instants greater than 10, suggesting, despite the small number of available samples, a greater robustness of the AI models when trained and used than for samples afferent to the last case.



(a) Number of passages of diagnostic train.



(b) Number of days for RUL.

Fig. 2: Distribution of measurement passages.

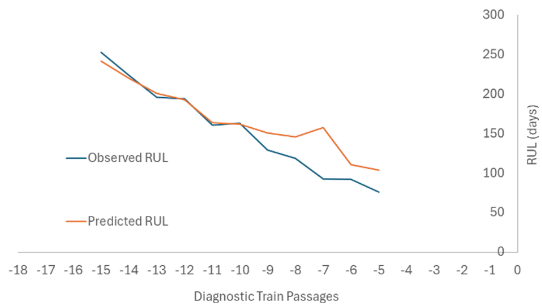


Fig. 3: Prediction performance as a function of available measurement obtained via diagnostic train passages.

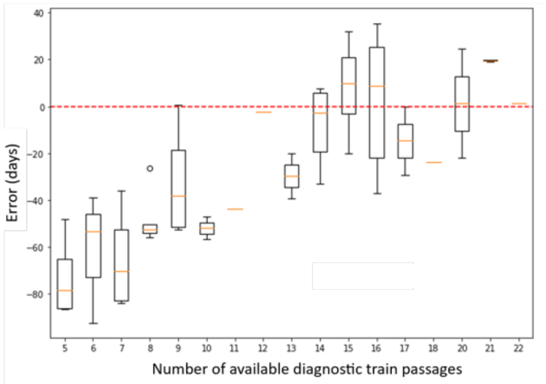


Fig. 4: Prediction error variation as a function of the number of available diagnostic train passages.

Predictor Variant		Model Performance	
Segment Length [m]	**Feature**	**MAE**	**RMSE**
600	Mean	21.18	26.96
	Median	29.47	32.65
	Kurtosis	28.31	33.28
200	Mean	24.97	27.01
	Median	22.56	23.61
	Kurtosis	22.78	28.53

V. Conclusion

This work addressed the challenge of automated diagnostic and predictive asset management for railway networks. To this end, an algorithm based on artificial intelligence techniques was developed. The proposed algorithm relied on diagnostic measures as inputs, and aimed to find a relation with the observed faults on the rail network. In operation mode, the algorithm predicts maintenance needs based on current measurements.

Due to the nature of the use case and of the available data samples, a Convolutional Neural Networks was used for the proposed model to autonomously learn how to extract information from the data. The model was trained with respect to the data measured by a real diagnostic train. In order to isolate conditions when the tracks start from their full functionality and progressively degrade toward the maintenance intervention, the measures from a major intervention to another have been considered. The results prove the feasibility of the proposed machine learning model for predicting the RUL of railways systems in a predictive maintenance perspective.

As future development, we aim to test and analyse the proposed algorithm by using different datasets. Moreover, it would be possible to optimize the parameters of the Convolutional Neural Network so to increase the performance of the algorithm, as well as to spot the more sensitive parameters that could affect the outcome.

References

[1] M. Binder, V. Mezhuyev, and M. Tschandl, "Predictive maintenance for railway domain: A systematic literature review," *IEEE Engineering Management Review*, vol. 51, no. 2, pp. 120–140, 2023.

[2] R. Macedo, R. Benmansour, A. Artiba, N. Mladenović, and D. Urošević, "Scheduling preventive railway maintenance activities with resource constraints," *Electronic Notes in Discrete Mathematics*, vol. 58, pp. 215–222, 2017.

[3] Z. Su, A. Jamshidi, A. Núñez, S. Baldi, and B. De Schutter, "Multi-level condition-based maintenance planning for railway infrastructures–a scenario-based chance-constrained approach," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 92–123, 2017.

[4] G. N. Bifulco, G. E. Cantarella, and F. Simonelli, "Design of signal setting and advanced traveler information systems," *Journal of Intelligent Transportation Systems*, vol. 18, no. 1, pp. 30–40, 2014.

[5] R. J. Marsal, "Large scale testing of rockfill materials," *Journal of the Soil Mechanics and Foundations Division*, vol. 93, no. 2, pp. 27–43, 1967.

[6] W. Salim, "Deformation and degradation aspects of ballast and constitutive modelling under cyclic loading," 2004.

[7] T. Dahlberg, "Railway track settlements-a literature review," *Report for the EU project SUPERTRACK*, vol. 463, 2004.

[8] P. McMahon, T. Zhang, and R. Dwight, "Requirements for big data adoption for railway asset management," *Ieee Access*, vol. 8, pp. 15 543–15 564, 2020.

[9] R. Tang, L. De Donato, N. Besinović, F. Flammini, R. M. Goverde, Z. Lin, R. Liu, T. Tang, V. Vittorini, and Z. Wang, "A literature review of artificial intelligence applications in railway systems," *Transportation Research Part C: Emerging Technologies*, vol. 140, p. 103679, 2022.

[10] W. Phusakulkajorn, A. Núñez, H. Wang, A. Jamshidi, A. Zoeteman, B. Ripke, R. Dollevoet, B. De Schutter, and Z. Li, "Artificial intelligence in railway infrastructure: current research, challenges, and future opportunities," *Intelligent Transportation Infrastructure*, vol. 2, p. liad016, 2023.

[11] A. Consilvio, J. Solís-Hernández, N. Jiménez-Redondo, P. Sanetti, F. Papa, and I. Mingolarra-Garaizar, "On applying machine learning and simulative approaches to railway asset management: The earthworks and track circuits case studies," *Sustainability*, vol. 12, no. 6, p. 2544, 2020.

[12] Z. M. Çınar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability*, vol. 12, no. 19, p. 8211, 2020.

[13] J. Xie, J. Huang, C. Zeng, S.-H. Jiang, and N. Podlich, "Systematic literature review on data-driven models for predictive maintenance of railway track: Implications in geotechnical engineering," *Geosciences*, vol. 10, no. 11, p. 425, 2020.

[14] Q. Wang, S. Bu, and Z. He, "Achieving predictive and proactive maintenance for high-speed railway power equipment with lstm-rnn," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6509–6517, 2020.

[15] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999–7019, 2021.