



## Rilievo delle irregolarità laterali del binario attraverso misure di accelerazioni laterali e spostamenti da bordo treno e algoritmi di Machine Learning

### *Detecting lateral track irregularities by onboard measurements of lateral acceleration and displacements and Machine Learning algorithms*

Nadia KAVIANI<sup>(\*)</sup>

Anders RØNNQUIST<sup>(\*\*)</sup>

Gunnstein Thomas FRØSETH<sup>(\*\*)</sup>

Albert LAU<sup>(\*\*\*)</sup>

Stefano RICCI<sup>(\*)</sup>

Luca RIZZETTO<sup>(\*)</sup>

(<https://www.medra.org/servlet/view?lang=it&doi=10.57597/IF.09.2024.ART.1>)

**Sommario** - Comprendere lo sviluppo dei difetti del binario, che possono avere un impatto significativo sulla sicurezza, è un compito cruciale per gli ingegneri ferroviari. L'analisi dei parametri della geometria è necessaria per monitorare in modo efficiente le condizioni del binario. La ricerca si è recentemente concentrata sul monitoraggio delle irregolarità della geometria del binario utilizzando i dati raccolti dai treni in servizio commerciale. La maggior parte della ricerca precedente si è concentrata sulle irregolarità verticali del binario, poiché le irregolarità laterali sono notevolmente più impegnative da derivare a causa del complesso moto relativo tra ruota e rotaia. Questo lavoro si concentra sulla ricerca della correlazione tra le accelerazioni laterali a bordo e le irregolarità laterali del binario, a partire dagli spostamenti laterali della ruota rispetto alla rotaia. Un algoritmo di *Machine Learning* (ML) supervisionato viene addestrato e testato con una serie di dati ottenuti tramite simulazione numerica in diverse condizioni operative. Tra gli algoritmi testati, la regressione *Random Forest* mostra il miglior risultato, seguita dalla regressione polinomiale. I risultati ottenuti sono una conferma della possibilità di rilevare le irregolarità laterali del binario mediante un sensore di visione computerizzato a bordo in grado di rilevare gli spostamenti laterali relativi tra ruota e rotaia. Inoltre, viene eseguita un'analisi di sen-

**Summary** - Understanding the development of track defects, which can have a significant impact on the safety of train operations, is one of the most crucial responsibilities for railway engineers. Analysing track geometry (TG) parameters is necessary to establish an efficient track condition monitoring approach. Research has recently focused on monitoring track geometry irregularities using data collected from trains in service. Most of the previous research has focused on vertical irregularity of track, since the lateral irregularity is considerably more challenging because of the complex relative wheel-rail motion. This research work focuses on finding the relationship between the lateral acceleration and lateral irregularities from the Lateral Displacement of the Wheel relative to the Rail (LDWR). A supervised Machine Learning (ML) model is trained and tested with a dataset composed of numerical simulation on a diverse set of operational conditions. Among the tested algorithms, Random Forest regression shows the best result and polynomial regression is on the second place. Sufficient results are achieved which shows that it is possible to detect the lateral irregularities with on board computer vision sensor which can detect the LDWR. Also, a sensitivity analysis is performed which shows that lateral displacement is the most important parameter in determining the lateral irregularity from vehicle monitoring. The result

<sup>(\*)</sup> Dipartimento Ingegneria Civile Edile e Ambientale (DICEA), Università La Sapienza di Roma, Via Eudossiana 18, 00184 Roma, Italia.

<sup>(\*\*)</sup> Dipartimento di Ingegneria Strutturale NTNU, Richard Birkelands vei 1A, 7034 Trondheim, Norvegia.

<sup>(\*\*\*)</sup> Dipartimento di Ingegneria Civile e Ambientale NTNU, Høgskoleringen 7a, 7491 Trondheim, Norvegia.

<sup>(\*)</sup> Department of Buildings and Environmental Engineering (DICEA), Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy.

<sup>(\*\*)</sup> Department of Structural Engineering, NTNU, Richard Birkelands vei 1A, 7034 Trondheim, Norway.

<sup>(\*\*\*)</sup> Department of Civil and Environmental Engineering, NTNU, Høgskoleringen 7a, 7491 Trondheim, Norway.

sibilità che mostra come questi rappresentino il parametro più importante per determinare le irregolarità laterali dal veicolo. I risultati di questo lavoro possono rappresentare un passo avanti verso l'impiego della manutenzione predittiva.

## 1. Introduzione

Riconoscere i difetti dei binari è un compito cruciale per garantire la sicurezza ferroviaria. L'analisi delle misure della geometria del binario è essenziale per sviluppare una strategia efficace di monitoraggio delle sue condizioni [1][2][3]. La norma europea EN13848-1 [4] definisce i seguenti parametri della geometria del binario da misurare per valutarne la qualità ai fini manutentivi: scartamento, livello trasversale, livello longitudinale, allineamento e sghembo [5]. Attraverso un'analisi statistica completa condotta durante un progetto di ricerca europeo [6], è emerso come il livello longitudinale e l'allineamento siano i principali parametri che influenzano le decisioni manutentive. Infatti, i parametri rimanenti presentano un tasso di degrado più lento o una robusta correlazione lineare con il livello longitudinale. La maggior parte dei gestori delle infrastrutture ferroviarie esegue il monitoraggio del binario utilizzando apparecchiature sofisticate montate su veicoli ferroviari diagnostici specializzati per misurarne i difetti della geometria [7]. Con queste tecnologie si misurano accuratamente i difetti del binario, ma il processo è costoso e richiede molto tempo. Ecco perché numerosi studi negli ultimi anni si sono concentrati sull'identificazione di tecniche innovative di monitoraggio delle condizioni del binario utilizzando misurazioni dinamiche da veicoli in servizio [8]. Oltre all'allineamento e allo scartamento, tutti i parametri della geometria del binario possono essere rilevati efficacemente utilizzando accelerometri. Sfortunatamente, la misurazione precisa dell'allineamento è piuttosto impegnativa e richiede l'uso di sensori ottici. Tuttavia, l'implementazione di tali sensori comporta un compromesso, poiché essi richiedono intense attività di manutenzione per garantirne la costante efficacia [9]. Studi precedenti hanno dimostrato che il livello longitudinale può essere ricostruito con precisione dalle accelerazioni verticali [8]. Tuttavia, non è lo stesso per l'allineamento. Pertanto, l'obiettivo di questo studio è identificare la relazione tra spostamenti laterali relativi ruota-rotaia e accelerazioni laterali della ruota, analizzando i dati registrati a bordo treno tramite algoritmi di apprendimento automatico per ricostruire l'allineamento del binario e rendere possibile in futuro l'impiego sistematico della manutenzione predittiva.

## 2. Esposizione del problema

Per una visione più chiara di questo problema, si fa riferimento alla Fig. 1(a). Nel piano orizzontale, nel punto P di ogni rotaia, che si trova 14 mm sotto il piano del ferro per il profilo di rotaia standard UIC 60E1, l'allineamento è definito come la deviazione laterale tra gli allineamenti

of this work could be a step toward the predictive maintenance as well.

## 1. Introduction

Recognizing track problems is a crucial responsibility for railway engineers to ensure safe train operations. Analysing the Track Geometry (TG) characteristics is essential for developing an effective track condition monitoring strategy [1][2][3]. TG parameters are defined as the indicators for track maintenance in the European railway standard EN13848-1 [4], namely track gauge, cross-level, longitudinal level, lateral alignment, and twist [5]. Through a comprehensive statistical analysis conducted during a European project [6], it was revealed that longitudinal level and lateral alignment emerge as the most crucial factors influencing maintenance decisions. Notably, the remaining parameters either exhibit a slower rate of degradation or demonstrate a robust linear correlation with the longitudinal level. Most of railway infrastructure managers perform track monitoring by using sophisticated equipment mounted on specialized diagnostic railway vehicles to measure geometry anomalies [7]. This technology measures the track defects accurately, but the process is costly and time-consuming. This is why numerous studies in recent years have concentrated on identifying innovative track condition monitoring techniques utilizing dynamic measurements from in-service vehicles [8].

Apart from lateral alignment and track gauge, all track geometry parameters can be effectively detected using accelerometers. Unfortunately, precise measurement of lateral alignment is quite challenging, necessitating the use of optical sensors. Nevertheless, the implementation of optical sensors comes with a trade-off, as they demand significant maintenance efforts to ensure their continued effectiveness [9].

Previous studies have shown that the longitudinal level can be accurately reconstructed from vertical accelerations [8]. However, it is not the same for lateral alignment. Hence, the objective of this study is to identify the relationship between lateral displacement of the wheel/rail, and lateral acceleration of the wheel by analysing on-board geometry monitoring data through machine learning algorithms to reconstruct the lateral alignment of the track and make it possible to reach to predictive maintenance in future.

## 2. Statement of problem

To have a clearer look on this issue, refer to Fig. 1 (a) In the horizontal plane, at point P on each rail, which is located 14 mm below the top of the railhead for the standard rail profile UIC 60E1, lateral alignment is defined as the lateral deviation between the actual and reference rail lines [4]. For lateral alignment to be estimated by accelerations, it is expected that the vehicle wheels would follow the impulse of lateral alignment in the lateral direction. The wheels do not, however, follow lateral alignment in the same manner as vertical alignment since the wheel has a freedom of movement in the lateral direction in a clearance between the wheel flange and rail head edge, shown

effettivi e di riferimento delle rotaie [4]. Affinché l'allineamento venga stimato tramite accelerazioni, ci si aspetta che le ruote del veicolo seguano l'impulso dell'allineamento nella direzione laterale. Tuttavia, le ruote non seguono l'allineamento laterale nello stesso modo dell'allineamento verticale, poiché la ruota ha un grado di libertà in direzione laterale pari alla distanza tra il bordino e il fianco della rotaia, mostrata nella Fig. 1(b), anche influenzato dal moto di serpeggio (vedi Fig. 1(c)). Pertanto, il problema è trovare una soluzione per rilevare le irregolarità laterali del binario, che in questo caso potrebbe essere la misura degli spostamenti laterali della ruota rispetto alla rotaia. Dopo aver risolto il problema del sistema di monitoraggio a bordo come illustrato in uno studio precedente [10], l'obiettivo principale è trovare un modello di apprendimento automatico per rilevare le irregolarità laterali tramite il monitoraggio da bordo e migliorare la correlazione tra allineamento e accelerazione laterale.

in Fig. 1(b), also affected by the hunting motion (see Fig. 1(c)). Therefore, the problem is to find a solution that could assess detecting the lateral irregularity, which in here could be detecting the lateral displacement of wheel to rail.

After solving the on-board monitoring issue which was mentioned in the previous study [10], the main objective is to find a Machine Learning model to detect the lateral track irregularity by on-board monitoring and solve the not solved issue of not having a good relationship between lateral alignment and lateral acceleration.

### 3. Problem description

TRUE et al. [11] conducted a theoretical investigation and found that determining lateral alignment only from vehicle accelerations is not feasible due to complex vehicle dynamics. To address this limitation, IN2SMART Project

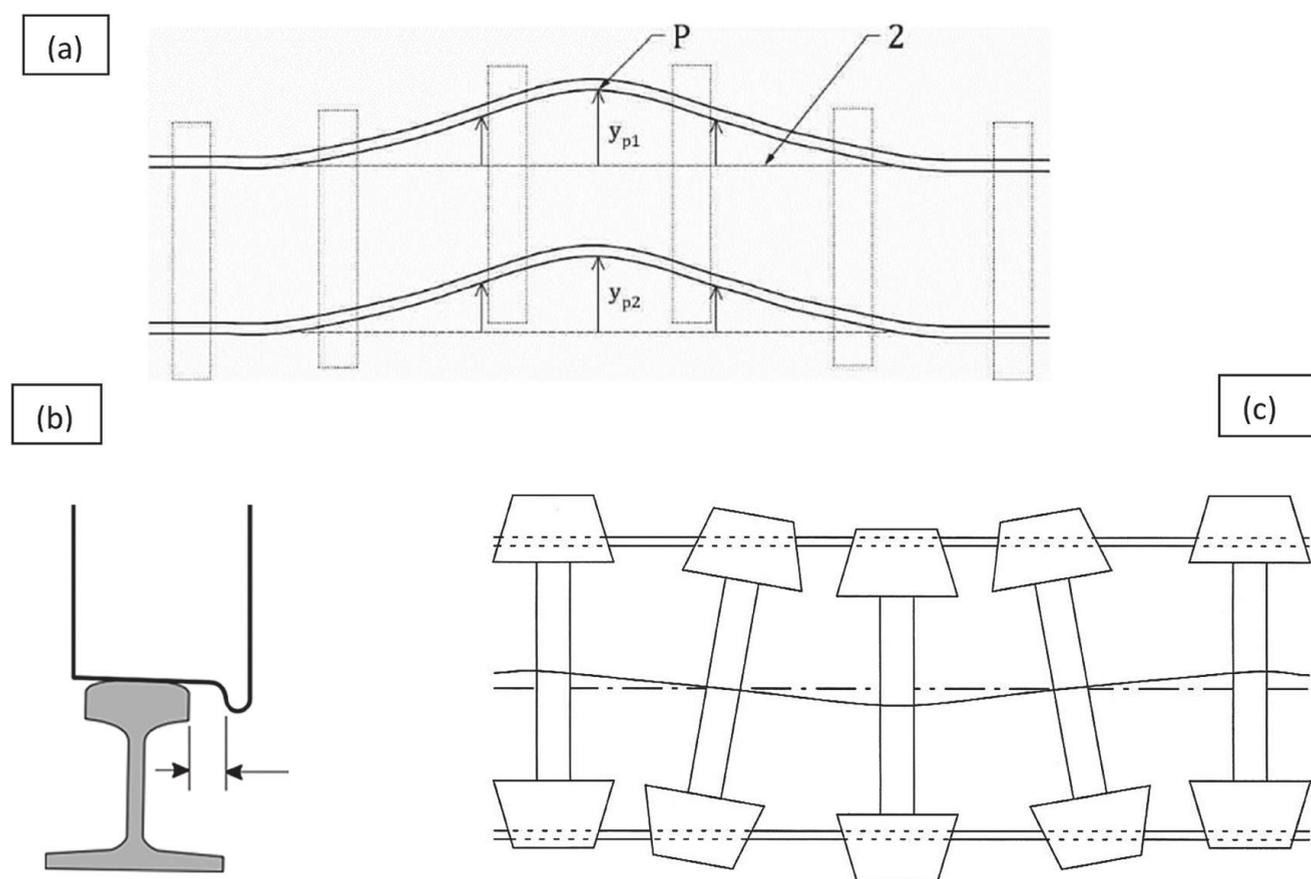


Figura 1 – (a) Allineamento laterale della rotaia sinistra  $y_{(P_1)}$  e della rotaia destra  $y_{(p_2)}$ , dove P indica il punto di riferimento della rotaia e 2 indica la linea di riferimento della rotaia [3]; (b) illustrazione della distanza tra la ruota e la rotaia [7]; (c) rappresentazione schematica del moto di serpeggio [7].

Figure 1 – (a) Lateral alignment of the left rail  $y_{(P_1)}$  and right rail  $y_{(p_2)}$ , where P denotes the rail reference point and "2" denotes the reference rail line [3]; (b) illustration of wheel/rail gauge clearance [7]; (c) schematic view of hunting motion [7].

### 3. Analisi della letteratura

TRUE *et al.* [11] hanno condotto un'indagine teorica e hanno scoperto che determinare l'allineamento solo dalle accelerazioni del veicolo non è possibile a causa della dinamica complessa del veicolo. Per superare il problema, RIPKE *et al.* [6] hanno impiegato una combinazione di misure di accelerazioni e un modello di simulazione dinamica multi-corpo (*Multi-Body Simulation* = MBS) del veicolo. Il loro approccio utilizzava un meccanismo di correzione dedicato, ciononostante i risultati non hanno soddisfatto le aspettative se confrontati con le misure di un sistema di misura della geometria del binario commerciale. ROSA *et al.* [12] hanno proposto un metodo alternativo basato su modelli che integrano MBS e un filtro di Kalman per stimare l'allineamento laterale. L'incapacità del modello di considerare l'usura del profilo della ruota, che ha un impatto considerevole sulla dinamica del veicolo, è un'importante carenza di questo metodo, così come la non linearità del problema. Pertanto, ROSA *et al.* [13] hanno tentato di utilizzare l'apprendimento automatico di tipo ML con i livelli trasversali e le accelerazioni laterali per rilevare le irregolarità laterali importanti del binario, ma con un'accuratezza inferiore al 90%, dunque non soddisfacente. Gli studi finora presentati hanno dimostrato che le sole accelerazioni laterali non possono essere utilizzate per determinare con accuratezza le irregolarità laterali del binario e che esse devono essere integrate con altri dati.

SHI *et al.* [10] hanno recentemente presentato un metodo per misurare gli spostamenti laterali in tempo reale e hanno suggerito che essi possano essere utili per stimare con accuratezza l'allineamento. La tecnica di misura utilizzata in questo metodo si basa su una telecamera stereoscopica ed è stata sviluppata come parte del progetto Assets4Rail [14][15]. Questo sistema di sensori è stato sviluppato specificamente per fornire come output gli spostamenti laterali relativi della ruota rispetto alla rotaia; a tal fine, le uscite video della telecamera stereo sono state analizzate utilizzando un algoritmo di elaborazione delle immagini. Questo metodo si prefigge di individuare una possibile soluzione al problema irrisolto di ricostruire correttamente l'allineamento del binario, combinando opportunamente le misure degli spostamenti laterali, ottenute tramite il sistema di sensori sopra citato, con le accelerazioni laterali rilevate da accelerometri a bordo del veicolo [6][10].

In questo articolo viene presentata una metodologia per determinare le irregolarità laterali del binario mediante l'utilizzo combinato delle misure degli spostamenti laterali della ruota rispetto alla rotaia, delle accelerazioni laterali della ruota e della conicità equivalente. Inoltre, viene eseguita un'analisi di sensibilità per identificare i parametri più importanti per il monitoraggio da bordo delle irregolarità laterali del binario.

Il seguente capitolo 4 descrive la metodologia proposta, inclusa l'illustrazione della simulazione numerica svolta e del metodo ML applicato ai risultati della simula-

[6] employed a combination of acceleration measurements and a Multi-Body dynamic Simulation (MBS) model of the vehicle. While their approach utilized a dedicated correction mechanism, the results fell short of expectations when compared to measurements from a commercial track geometry inspection system. ROSA *et al.* [12] proposed an alternative model-based method that integrated MBS and a Kalman filter to estimate lateral alignment. The model's inability to consider wheel profile wear, which has a considerable impact on vehicle dynamics, is an important weakness of this technique, as well as non-linearity of the problem. Therefore, ROSA *et al.* [13] attempted to use machine learning (ML) with cross-level and lateral acceleration information to detect large track lateral irregularities, but the accuracy was less than 90% and not satisfying.

The studies presented above have shown that lateral accelerations alone cannot be used to accurately determine the lateral irregularity and that accelerations must be supplemented by other proper response quantities.

SHI *et al.* [10] recently presented a method to measure LDWR in a real-time and suggested that LDWR can be useful to accurately estimate lateral alignment. The measurement technique utilized in this method relies on a stereo camera, which was developed as part of the Assets4Rail project [14] [15]. This sensor system was developed specifically to provide the LDWR as an output, and then the video outputs of the stereo camera were analyzed using the image processing algorithm. It offers a possible answer to the unsolved problem of properly reconstructing alignment on in-service cars by fusing this suggested method with accelerometers [6][10].

In the current work, a methodology for determining lateral irregularities by incorporating LDWR, lateral accelerations, and conicity is presented. Furthermore, a sensitivity analysis to identify the most crucial parameters for on-board monitoring of lateral irregularity is performed.

Section 4 describes the methodology, including the explanation of the numerical simulation and the ML method applied to the simulation results. The findings using the ML method are presented as the results of this paper.

## 4. Methodology

### 4.1. Estimating the lateral irregularity of track with onboard response measurements

The methodology employed in this study aims to estimate lateral irregularity by utilizing measurements of LDWR and lateral acceleration from a train, in conjunction with a trained ML model. The rationale behind choosing ML models lies in the inverse problem which this study seeks to address and using system responses such as acceleration, LDWR, and conicity to estimate lateral irregularity accurately is the final goal. Machine learning regression methodology is proposed as it has demonstrated remarkable efficiency in handling various forms of non-linearity that may be present in the data, making it well-suited for tackling the complexities of the lateral irregularity estimation problem.

zione. Vengono, quindi, presentati i risultati ottenuti utilizzando il metodo ML.

**4. Metodologia**

**4.1. Stima delle irregolarità laterali della rotaia con misurazioni da bordo treno**

La metodologia impiegata in questo studio mira a stimare le irregolarità laterali del binario utilizzando le misurazioni di spostamenti e accelerazioni laterali effettuate da bordo treno, assieme a un modello di apprendimento automatico (ML) addestrato (Fig. 2). La logica alla base della scelta dei modelli ML risiede nel problema inverso che questo studio cerca di affrontare e l'utilizzo di risposte di sistema come accelerazioni, spostamenti laterali relativi ruota-rotaia e conicità per stimare accuratamente le irregolarità laterali rappresenta l'obiettivo finale. La metodologia di regressione di apprendimento automatico è proposta in quanto ha dimostrato una notevole efficienza nella gestione di varie forme di non linearità che possono essere presenti nei dati, rendendola adatta per affrontare le complessità del problema di stima delle irregolarità laterali. Idealmente, per ottenere risultati accurati, sarebbe richiesta una conoscenza perfetta del sistema e del processo, il che comporta la misurazione di tutti i parametri di input e risposta [16]. Tuttavia, limitazioni pratiche come vincoli economici e tecnologici rendono impossibile raccogliere dati così completi. Per superare questa problematica, è stata utilizzata la simulazione numerica per valutare il metodo proposto, sviluppare algoritmi ed eseguire uno studio di sensibilità per identificare le variabili chiave per future misurazioni in campo. Il metodo suggerito per la stima delle irregolarità laterali è quindi quello di misurare spostamenti e accelerazioni su un binario con irregolarità laterali note per costruire un modello ML, quindi misurare i dati di input su qualsiasi altro binario e applicare il modello ML addestrato per stimare le irregolarità del binario.

*Ideally, to achieve accurate results, perfect knowledge of both the system and the process, which entails measuring all input and response parameters would be required [16]. However, practical limitations such as economic and technological constraints make it infeasible to gather such comprehensive data. To overcome these challenges, numerical simulation was used to evaluate the proposed method, develop algorithms, and perform a sensitivity study to identify key variables for future real-world measurements.*

*The suggested method for lateral irregularity estimation is therefore to measure LDWR and acceleration at a track with known lateral irregularity to build a ML model, then run the train with measurements on any other track and apply the trained ML model to estimate track irregularity.*

*Hence, in this study, the focus is on detecting track faults using various machine learning models. To achieve this, relevant features have been carefully selected. The methodology follows a systematic approach, commencing with the simulation phase. Subsequently, the steps of the methodology are outlined in the following sections.*

*The simulation was performed using Gensys multibody simulation software on two standard tracks: Czech high-speed test track in Velim (ST1) and Polish lower-speed test track in Zmigrod (ST2), both have various curves and inclinations. The output of this step would be the features of the model described in Fig. 2.*

*Different algorithms which are used in this study are random forest regression, polynomial regression, and support vector machine regression, to predict lateral irregularity by using numerical simulated lateral acceleration of the train bogie and lateral displacement of the leading wheelset.*

**4.2. Numerical simulation and case study**

*To simulate the vehicle-track dynamic behaviour, and analyse if the proposed idea is working well, numerical sim-*

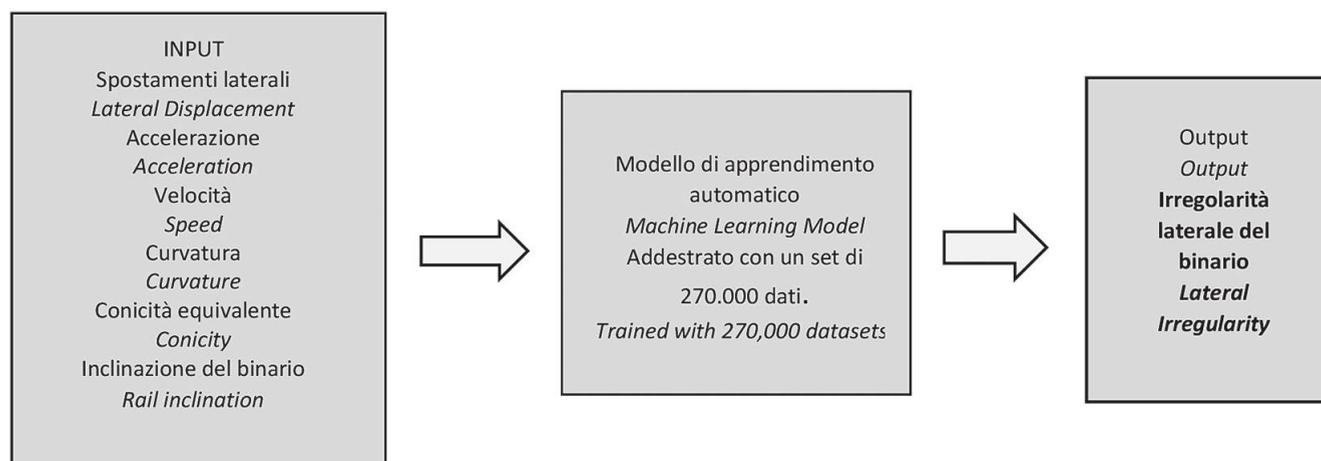


Figura 2 – Input e output del modello di apprendimento automatico.  
Figure 2 – Input-Output for Machine Learning Model.

Pertanto, in questo studio, l'attenzione è rivolta al rilevamento di difetti laterali del binario utilizzando vari modelli di apprendimento automatico. Per raggiungere questo obiettivo, sono state attentamente selezionate le grandezze da prendere in considerazione nel modello. La metodologia segue un approccio sistematico, a partire dalla fase di simulazione, con i passaggi che vengono illustrati nei paragrafi seguenti.

La simulazione è stata eseguita utilizzando il software di simulazione *Multi-body Gensys®* su due binari standard: il binario di prova ad alta velocità polacco di Zmigrod (ST2), entrambi con varie curve e considerando diverse inclinazioni di posa delle rotaie (Fig. 3). L'output delle simulazioni multicorpo sono le grandezze di input del modello di apprendimento automatico, elencate nella colonna di sinistra della Fig. 2.

I diversi algoritmi utilizzati in questo studio sono la regressione *Random Forest*, la regressione polinomiale e la *Support Vector Machine Regression*. Tali algoritmi sono stati utilizzati per prevedere le irregolarità laterali del binario a partire dai valori dell'accelerazione laterale del carrello e degli spostamenti laterali della sala montata di testa, ottenuti mediante le simulazioni numeriche.

**4.2. Simulazione numerica e caso di studio**

Per simulare il comportamento dinamico veicolo-via e verificare il funzionamento dell'idea proposta, sono state eseguite simulazioni numeriche con il software *Gensys®*. Nelle simulazioni è stato viaggiare su due diversi binari standard un veicolo composto da una cassa con due carrelli e quattro assili. La cassa, i carrelli e gli assili sono accoppiati con molle e smorzatori. Le caratteristiche del veicolo sono riportate in [17].

Le caratteristiche dei binari standard sono le seguenti:

- ST1 – Binario di prova ad alta velocità a Velim (Cechia), che include 2 curve, con una lunghezza totale del tracciato di 13.276 m.

*ulations have been carried out by the software Gensys®. A train running on two different Standard Tracks (ST) is used in this work. The train has one car body with two bogies and 4 wheelsets. Car body, bogies and wheelsets are rigid masses coupled by spring and dampers. The characteristic of the train can be found in [17].*

*The characteristics of the standard tracks (ST) are as follow and seen in Fig. 3.*

- ST1 – Czech high speed test track in Velim, including 2 curves, with a total track length 13,276 m.
- ST2 – Polish lower speed test track in Zmigrod, including several curves, with a length of test track. 7,725 m.

*These tracks were chosen due to their availability of access to their characteristics and well layout design of operation parameter [18][19]. In the simulation software, the lateral irregularity of the track was modelled using Swedish standard irregularity on the first quality level (QN1) based on the UIC 518.2009 [20] and EN 14363:2005 [21], which is called K0 for this work. Three distinct speeds were used for each track during the simulation test: ST1 speeds were 130, 160, and 210 km/h, and ST2 speeds were 50, 80, and 120 km/h. Different inclinations were used during the simulations, including 1:20, 1:30, and 1:40. Lateral irregularity was the only geometry parameter considered in the study, and 18 simulations were conducted, which could be seen in Tab. 1.*

*The Swedish wheel profile defined in EN S1002 T32.5, and the rail profile was 60 E1 with with the equivalent conicity calculated using the Gensys® software.. It affects the lateral stability and steering of the vehicle, with higher conicity leading to stronger self-centering forces and influencing ride quality and safety. The focus among the results of the simulation, is just the lateral acceleration of the wheel, and LDWR which could be used for the Machine Learning model inputs, with the goal to find the relationship between lateral track irregularities and the lateral displacement.*

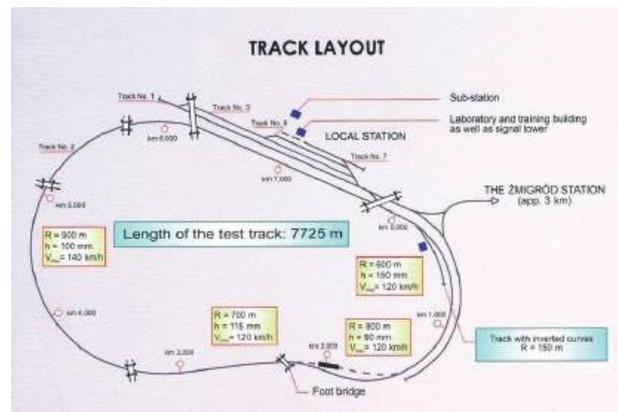


Figura 3 – Binari standard: circuito ceco di Velim (sinistra) e circuito polacco di Zmigrod (destra) [18][19].  
 Figure 3 – Standard Tracks: Czech Velim track (left), Polish Zmigrod track (right)[18][19].

Tabella 1 – Table 1

Simulazione Gensys®  
Gensys® Simulation

| Scenario<br>Scenario | Tipo di binario<br>Track type | Velocità (km/h)<br>Speed (km/h) | Inclinazione<br>Inclination | Tipo di irregolarità<br>Irregularity type |
|----------------------|-------------------------------|---------------------------------|-----------------------------|---|
| 1                    | ST2                           | 50                              | 1:20                        | K0  |
| 1                    | ST2                           | 50                              | 1:20                        | K0  |
| 2                    | ST2                           | 50                              | 1:30                        | K0  |
| 3                    | ST2                           | 50                              | 1:40                        | K0  |
| 4                    | ST2                           | 80                              | 1:20                        | K0  |
| 5                    | ST2                           | 80                              | 1:30                        | K0  |
| 6                    | ST2                           | 80                              | 1:40                        | K0  |
| 7                    | ST2                           | 120                             | 1:20                        | K0  |
| 8                    | ST2                           | 120                             | 1:30                        | K0  |
| 9                    | ST2                           | 120                             | 1:40                        | K0  |
| 10                   | ST1                           | 130                             | 1:20                        | K0  |
| 11                   | ST1                           | 130                             | 1:30                        | K0  |
| 12                   | ST1                           | 130                             | 1:40                        | K0  |
| 13                   | ST1                           | 160                             | 1:20                        | K0  |
| 14                   | ST1                           | 160                             | 1:30                        | K0  |
| 15                   | ST1                           | 160                             | 1:40                        | K0  |
| 16                   | ST1                           | 210                             | 1:20                        | K0  |
| 17                   | ST1                           | 210                             | 1:30                        | K0  |
| 18                   | ST1                           | 210                             | 1:40                        | K0  |

- ST2 – Binario di prova a bassa velocità di Zmigrod (Polonia), che include diverse curve, con una lunghezza totale del tracciato di 7.725 m.

Questi binari sono stati scelti per la disponibilità dei loro dati caratteristici e per la progettazione del layout [18][19]. Nel software di simulazione, l'irregolarità laterale del binario è stata modellata utilizzando le irregolarità standard svedesi sul primo livello di qualità (QN1) in base agli standard UIC 518.2009 [20] ed EN 14363:2005 [21], e viene denominata K0 nel presente lavoro. Sono state utilizzate tre velocità distinte per ciascun binario durante i test di simulazione: le velocità in ST1 sono state di 130, 160 e 210 km/h e le velocità in ST2 sono state di 50, 80 e 120 km/h. Durante le simulazioni sono state anche utilizzate diverse inclinazioni della rotaia: 1:20, 1:30 e 1:40. L'unico parametro geometrico preso in considerazione nello studio è stata l'irregolarità laterale e sono state condotte 18 simulazioni, come riportato nella Tab. 1.

Il profilo di ruota utilizzato nelle simulazioni era quello svedese definito nella EN S1002 T32.5 e il profilo della rotaia era il 60 E1 con la conicità equivalente calcolata per mezzo del software Gensys®. La conicità influisce sulla stabilità laterale e il comportamento in curva del veicolo; un incremento della conicità comporta forze autocentranti maggiori e influenza la qualità di marcia e la sicurezza. Tra gli output della simulazione ci si è focalizzati sull'accelerazione laterale della ruota e sullo spostamento laterale ruota-rotaia, che sono stati successivamente utilizzati

**4.3. Regression models by Machine Learning approach**

*Recently, the capability of ML models to monitor the health of structures and track geometry has been proven and is in a high interest of researchers and industry. As it was mentioned before, it is a powerful method since it is capable to handle the non-linearity of the problem.*

*Regression methods are used in this work to investigate the relationship between the variables and to confirm the simulation results. Regression is a technique which helps in examining the relationship between independent features or variables and a dependent feature, outcome, or variable. To ensure that altering one algorithmic parameter does not drastically change the results of the entire model, it is intended to start with simple regression algorithms that can support related case studies without relying on many algorithmic parameters.*

*For this project, supervised learning was chosen since it is a machine learning task of establishing a function that maps an input to an output based on example input-output pairs [22]. Regression entails choosing the best model to suit the provided data set and then using the model to make predictions in the future [23][24].*

*In different situations, various regression models are utilized, and their appropriateness depends on the type of data and the relationships between the variables. For instance, when dealing with multiple variables showing a polynomial relationship, the preferred choice is a Polynomial Regression*

come input del modello di Machine Learning, con l'obiettivo di trovare una relazione tra le irregolarità laterali del binario e lo spostamento laterale relativo della ruota rispetto alla rotaia

### 4.3. Modelli di regressione mediante approccio di apprendimento automatico

Idi recente, la capacità dei modelli ML di monitorare lo stato di salute delle strutture e la geometria del binario è stato dimostrato essere di grande interesse per la ricerca e l'industria ferroviaria. Come accennato in precedenza, si tratta di un metodo potente poiché è in grado di gestire la non linearità del problema.

In questo lavoro vengono utilizzati metodi di regressione per indagare la relazione tra le variabili e confermare i risultati della simulazione. La regressione è una tecnica che aiuta a esaminare la relazione tra caratteristiche o variabili indipendenti e una caratteristica, risultato o variabile dipendente. Per garantire che la modifica di un parametro dell'algoritmo non modifichi drasticamente i risultati dell'intero modello, si inizia con semplici algoritmi di regressione in grado di supportare casi di studio correlati senza basarsi su molti parametri algoritmici.

Per questo progetto, è stato scelto l'apprendimento supervisionato, attività di apprendimento automatico per stabilire una funzione che mappi input e output in base a coppie di esempio [22]. La regressione comporta la scelta del modello migliore per adattarsi all'insieme di dati fornito e quindi l'utilizzo del modello per fare previsioni in futuro [23] [24]. In diverse situazioni, vengono utilizzati vari modelli di regressione, e la loro appropriatezza dipende dal tipo di dati e dalle relazioni tra le variabili. Ad esempio, quando si ha a che fare con più variabili che mostrano una relazione polinomiale, la scelta preferita è un modello di regressione di questo tipo. D'altro canto, per i dati che mostrano una relazione lineare tra due variabili, l'opzione più adatta è un modello di regressione di questo secondo tipo [25]. Gli algoritmi utilizzati in questo studio sono la regressione *Random Forest*, la regressione polinomiale e la *Support Vector Machine Regression*, per prevedere le irregolarità laterali del binario utilizzando i valori simulati delle accelerazioni laterali del carrello e degli spostamenti laterali del primo assile.

#### 4.3.1. Analisi dei dati

Nel processo di training (addestramento) del modello, quattro parametri vengono impiegati come input per l'apprendimento supervisionato. Questi parametri sono le accelerazioni e gli spostamenti laterali della ruota, la conicità e le irregolarità laterali della rotaia. Quest'ultime sono utilizzate durante l'addestramento in quanto fungono da variabile obiettivo per l'approccio di apprendimento supervisionato.

Durante la fase di addestramento, il modello impara a rilevare irregolarità laterali utilizzando i dati forniti di ac-

*Model.* On the other hand, for data exhibiting a linear relationship between two variables, the most suitable option is a *Linear Regression Model* [25].

*Different algorithms which are used in this study are random forest regression, polynomial regression, and support vector machine regression, to predict lateral irregularity by using numerical simulated lateral acceleration of the train bogie and lateral displacement of the leading wheelset.*

#### 4.3.1 Data management:

*In the training process of the model, four parameters are employed as inputs for supervised learning. These parameters include the lateral acceleration of the wheel, LDWR, conicity, and the lateral irregularity of the track. Among these, the lateral irregularity is used during training as it serves as the target variable for the supervised learning approach.*

*Through the training phase, the model learns how to detect lateral irregularities by utilizing the provided lateral acceleration and LDWR data. The relationship between these input parameters and the target variable (lateral irregularity) is learned and captured by the model, allowing it to later predict or detect lateral irregularities solely based on the observed lateral acceleration and LDWR without directly using the lateral irregularity in the prediction process. This demonstrates the model's ability to generalize and identify lateral track irregularities based on the learned patterns from the training data.*

*Several attempts were made at the train-test, and it was found that changes in inclination could significantly affect the results if the model was not trained and tested on the same inclinations. To address this, three different groups were created based on the different inclinations, and the sample model was selected based on the highest accuracy of each group. Each sample model was then trained using the full dataset of that group and tested on other case studies with the same inclination.*

*The sample model chosen for analysis is the standard track 1 located in the Czech Republic. This model was trained at a speed of 130 km/h, considering three distinct inclinations: 1:20, 1:30, and 1:40. To assess and determine the most effective method for the stated objective, three regression techniques were employed: Random Forest Regression, Polynomial Regression, and Support Vector Machine Regressor.*

*During the testing phase, the track was divided into various sections, including the whole track, straight track segments, and only the curved sections. This division was implemented to gain deeper insights into the model's performance and finding the reason of some low accuracies by checking each predictive accuracy.*

*By evaluating the results on different track sections, the study aims to identify the strengths and weaknesses of each regression model, leading to the selection of the most suitable and reliable method for accurately predicting lateral irregularities based on the given dataset.*

celerazioni e spostamenti laterali. La relazione tra questi parametri di input e la variabile target (irregolarità laterali) viene appresa e catturata dal modello, consentendogli in seguito di prevedere o rilevare irregolarità laterali basandosi esclusivamente sulle accelerazioni e gli spostamenti laterali osservati, senza utilizzare direttamente le irregolarità laterali nel processo di previsione. Ciò dimostra la capacità del modello di generalizzare e identificare le irregolarità laterali del binario in base all'apprendimento dai dati di training.

Sono stati effettuati diversi test e si è scoperto che i cambiamenti dell'inclinazione di posa delle rotaie potrebbero influenzare significativamente i risultati se il modello non fosse stato addestrato e testato sui medesimi valori di inclinazione. Per risolvere questo problema, sono stati creati tre gruppi diversi in base alle diverse inclinazioni e il modello campione è stato selezionato in base alla massima accuratezza di ciascun gruppo. Ogni modello campione è stato quindi addestrato utilizzando il set di dati completo di quel gruppo e testato su altri casi di studio con lo stesso valore di inclinazione.

Il modello campione scelto per l'analisi è il binario ST1. Questo modello è stato addestrato a una velocità di 130 km/h, considerando tre inclinazioni distinte: 1:20, 1:30 e 1:40. Per valutare e determinare il metodo più efficace per l'obiettivo dichiarato, sono state impiegate le tecniche di regressione *Random Forest Regression*, regressione polinomiale e *Support Vector Machine Regression*.

Durante la fase di test, il tracciato è stato diviso in varie sezioni, tra cui l'intero tracciato, i tratti rettilinei e le curve. Questa divisione è stata implementata per ottenere informazioni più approfondite sulle prestazioni del modello e trovare il motivo di alcune basse accuratezze controllando ogni valore di accuratezza predittiva. Valutando i risultati su diverse sezioni di binario, lo studio mira a identificare i punti di forza e di debolezza di ciascun modello di regressione, portando alla selezione del metodo più adatto e affidabile per prevedere accuratamente le irregolarità laterali in base al set di dati fornito. Dopo aver applicato il modello su diversi casi di studio, gli errori sono stati calcolati e classificati dal migliore al peggiore, facendo riferimento alle due metriche seguenti.

#### 4.3.2. Misure di prestazione

La valutazione delle prestazioni del modello ha comportato il calcolo dei valori *Root Mean Squared Error* (RMSE) e *R-squared* (R<sup>2</sup>). Questi due indicatori sono comunemente utilizzati dagli statistici per valutare l'adattamento di un modello di regressione a un set di dati. Il RMSE misura la differenza media tra i valori previsti e quelli effettivi nel set di dati e non è rapportato a nessun valore. D'altro canto, R<sup>2</sup>, noto anche come coefficiente di determinazione, varia tra 0 e 1, dove i valori più vicini a 1 indicano un migliore adattamento del modello al set di dati. L'obiettivo era ottenere un RMSE basso e un R<sup>2</sup> alto per dimostrare l'efficacia del modello [26]. Essi sono calcolati come:

*After applying the model on different case studies, errors were calculated and identified as based on finding best and worst which would be explained by referring to the two following metrics:*

#### 4.3.2. Performance metrics

*The evaluation of the model's performance involved calculating the Root Mean Squared Error (RMSE) and R-squared (R<sup>2</sup>) values. These two indicators are commonly used by statisticians to assess the fit of a regression model to a dataset. The RMSE measures the average difference between the predicted and actual values in the dataset and is not scaled to any value. On the other hand, R<sup>2</sup>, also known as the coefficient of determination, ranges between 0 and 1, where values closer to 1 indicate a better fit of the model to the dataset. The aim was to obtain a low RMSE and a high R<sup>2</sup> to demonstrate the effectiveness of the model [26].*

*They are calculated as:*

$$RMSE = \sqrt{\sum(P_i - O_i)^2/n}$$

*where:*

- *P<sub>i</sub> is the predicted value for the i<sup>th</sup> observation;*
- *O<sub>i</sub> is the observed value for the i<sup>th</sup> observation;*
- *n is the sample size;*
- *R<sup>2</sup> = 1 - (RSS/TSS);*
- *RSS (Residual Sum of Squares): represents the sum of the squared differences between the observed values and the predicted values;*
- *TSS (Total Sum of Squares): represents the total sum of the squared differences between the observed values and their mean.*

#### 4.3.3. Data aggregation to develop the model

*Through comprehensive data analysis, certain parameters were chosen among the simulation outcomes that exhibit good influence on the model's predictive capabilities. To create a generalized model, all these important parameters were included as input features. Then, a 70/30 data split was used to train and evaluate this model. Several parameters and track variables are included in the model, including LD-WR, lateral acceleration, speed, inclination, curvature, and conicity. Conicity was used as an input characteristic in the model even though it cannot be measured. The model was put to the test twice, once using conicity as input and once without, to evaluate its influence and the performance of the model. The study's subsequent chapter details the outcomes of these tests, including the distribution of errors discovered during the evaluation process.*

#### 4.3.4. Sensitivity analysis

*As it was mentioned earlier, ML methods have been widely used, especially to solve high-dimensional and non-linear problems [27][28]. However, handling high-dimensional data poses difficulties for decision-making and data analysis. Feature Selection (FS) techniques have been developed*

$$RMSE = \sqrt{\sum(P_i - O_i)^2/n}$$

In cui:

- $\sum$  è il simbolo di sommatoria;
- $P_i$  è il valore previsto per la i-esima osservazione;
- $O_i$  è il valore osservato per la i-esima osservazione;
- $n$  è la dimensione del campione;
- $R^2 = 1 - (RSS/TSS)$ ;
- **RSS (Residual Sum of Squares)**: rappresenta la somma delle differenze al quadrato tra i valori osservati e i valori previsti;
- **TSS (Total Sum of Squares)**: rappresenta la somma totale delle differenze al quadrato tra i valori osservati e la loro media.

#### 4.3.3. Raccolta dei dati per sviluppare il modello

Attraverso un'analisi completa dei dati, sono stati scelti alcuni parametri tra i risultati della simulazione che mostrano una buona influenza sulle capacità predittive del modello. Per creare un modello generalizzato, tutti questi parametri sono stati inclusi come caratteristiche di input. Quindi, è stata utilizzata una suddivisione dei dati 70/30 per valutare il modello. Diversi parametri del binario sono compresi nel modello, tra cui spostamenti e accelerazioni laterali, velocità, inclinazione della rotaia, curvatura del tracciato e conicità della ruota. La conicità è stata utilizzata come caratteristica di input nel modello anche se non può essere misurata nella realtà. Il modello è stato messo alla prova due volte, una volta utilizzando la conicità come input e una volta senza, al fine di valutarne l'influenza e le prestazioni del modello. Il paragrafo successivo descrive in dettaglio i risultati di questi test, inclusa la distribuzione degli errori scoperti durante il processo di valutazione.

#### 4.3.4 Analisi di sensibilità

Come accennato in precedenza, i metodi di apprendimento automatico (ML) sono stati ampiamente utilizzati, in particolare per risolvere problemi ad alta complessità e non lineari [27][28]. Tuttavia, la gestione di dati ad alta complessità (dati che presentano un gran numero di caratteristiche o attributi) pone difficoltà per il processo decisionale e l'analisi dei dati stessi. Le tecniche di selezione delle caratteristiche (Feature Selection = FS) sono state sviluppate con l'obiettivo di elaborare in modo efficiente dati ad alta complessità per affrontare questo problema. Per migliorare la correttezza del modello, il metodo FS cerca il miglior set di caratteristiche possibile, che fornisca la maggior parte delle informazioni [29]. Sulla base di un set di parametri, CAI et al. [30] hanno categorizzato gli approcci FS in più gruppi. Le tre categorie generali dei metodi FS sono i modelli di *filtering*, *wrapper* ed *embedding*, che si distinguono per le rispettive connessioni agli approcci di apprendimento. In questo lavoro, è stata scelta l'importanza della caratteristica di permutazione per l'analisi di sensibilità. Quando i dati sono tabellari, un metodo di ispezione del modello chiamato importanza

with the goal of efficiently processing high-dimensional data in order to address this problem. To improve model correctness, the FS method looks for the best possible feature set that yields the most information [29]. Based on a variety of parameters, Cai et al. [30] categorized FS approaches into multiple groups. The three general categories of FS methods are filter, wrapper, and embedding models, which are distinguished by their respective connections to learning approaches.

In this work, the importance of the Permutation feature was chosen for sensitivity analysis. When the data is tabular, a model inspection method called permutation feature importance can be used to any fitted estimator. This is particularly helpful for unclear or non-linear estimators. The reduction in a model score resulting from a single feature value being shuffled at random is known as the permutation feature significance [31]. As a result, it can evaluate the relative impact of each quantity on the issue; the outcomes are displayed in the following paragraph.

#### 4.3.5. Outline of the permutation importance algorithm

As it was mentioned in the previous section, the permutation importance algorithm is a method used to assess the importance of each feature in a predictive model. This technique involves evaluating the model's performance after systematically shuffling each feature to break its association with the target variable. The decrease in the model's performance indicates the importance of the feature: the more significant the drop, the more important the feature is [31]. The following section details the permutation importance algorithm, explaining the steps involved and presenting the formula used to calculate the importance of each feature:

- **Inputs:** fitted predictive model  $m$ , tabular dataset (training or validation)  $D$ .
- Compute the reference score  $s$  of the model  $m$  on data  $D$  (for instance the accuracy for a classifier or the  $R^2$  for a regressor).
- For each feature  $j$  (column of  $D$ ):

For each repetition  $k$ , which is counting the permutations of feature  $f_j$  for which the performance metric  $s_{(k,j)}$  is calculated and then averaged to determine the feature's importance:

- Randomly shuffle column  $j$  of dataset  $D$  to generate a corrupted version of the data named  $\tilde{D}_{k,j}$ .
- Compute the score  $s_{(k,j)}$  of model  $m$  on corrupted data  $\tilde{D}_{k,j}$ .
- Compute importance  $i_{(k,j)}$  for feature  $f_j$  defined as [26]:

$$i_j = s - \frac{1}{k} \sum_{k=1}^k s_{k,j}$$

In this formula:

- $i_j$  represents the importance score for feature  $f_j$ ;

della caratteristica di permutazione può essere utilizzato per qualsiasi stimatore adattato. Ciò è particolarmente utile per stimatori poco chiari o non lineari. Il fatto che i valori di una singola caratteristica vengano mescolati in modo casuale, per osservare la conseguente degradazione del punteggio del modello, è noto come significatività della caratteristica di permutazione [31]. Di conseguenza, il metodo consente di valutare l'impatto relativo di ciascuna caratteristica sul problema; i risultati dell'applicazione del metodo sono presentati nel paragrafo seguente.

#### 4.3.5. Schema dell'algoritmo di importanza della caratteristica di permutazione

Come accennato nella sezione precedente, l'algoritmo di importanza della caratteristica di permutazione è un metodo utilizzato per valutare l'importanza di ogni caratteristica in un modello predittivo. Questa tecnica comporta la valutazione delle prestazioni del modello dopo aver sistematicamente mescolato ogni caratteristica per interromperne l'associazione con la variabile target. La diminuzione delle prestazioni del modello indica l'importanza della caratteristica: più il calo è significativo, più importante è la caratteristica [31]. La sezione seguente descrive in dettaglio l'algoritmo di importanza della caratteristica di permutazione, spiegandone i passaggi e presentando le formule utilizzate per calcolare l'importanza di ogni caratteristica:

- Input: modello predittivo adattato  $m$  e set di dati tabellari  $D$  di training o validazione.
- Calcolo del punteggio di riferimento  $s$  del modello  $m$  sui dati  $D$  (ad esempio l'accuratezza per un classificatore o la  $R^2$  per un regressore).
- Per ogni caratteristica  $j$  (colonna di  $D$ ):
  - Per ogni ripetizione  $k$ , che conta le permutazioni della caratteristica  $f_j$  per la quale si calcola  $s_{(k,j)}$  e se ne fa la media per determinare l'importanza della caratteristica:
    - Colonna  $j$  del dataset  $D$  distribuita casualmente per generare una versione alterata dei dati chiamata  $\tilde{D}_{k,j}$ .
    - Calcolo del punteggio  $s_{(k,j)}$  del modello  $m$  sui dati  $\tilde{D}_{k,j}$ .
  - Calcolo dell'importanza  $i_{(k,j)}$  per la caratteristica  $f_j$  definita come in [26]:

$$i_j = s - \frac{1}{k} \sum_{k=1}^k s_{k,j}$$

In questa formula:

- $i_j$  rappresenta il punteggio di importanza della caratteristica  $f_j$ ;
- $s$  è una metrica di base (ad esempio: accuratezza, RMSE) calcolata utilizzando i dati originali;
- $s_{(k,j)}$  è la metrica calcolata dopo aver permutato i valori della caratteristica  $f_j$  nel  $k$ -esimo caso;

- $s$  is a baseline metric (e.g., accuracy, RMSE) computed using the original data;
- $s_{(k,j)}$  is the metric computed after permuting the values of feature  $f_j$  in the  $k$ -th instance;
- The summation  $\sum_{k=1}^k s_{k,j}$  indicates the average metric over  $k$  permutations;
- $k$  refers to the number of permutations performed for the feature; essentially, it counts the number of times the feature  $f_j$  is permuted and the metric  $s_{(k,j)}$  is calculated.

## 5. Results and discussion

### 5.1. Influence of track inclination on estimation of lateral track irregularity

While analysing the data output from the Gensys® simulation for model creation and development, a notable finding emerged as an initial result. This interesting finding revolves around the impact of rail inclination on the model.

Fig. 4 illustrates the initial validation attempt, which resulted in failure. This outcome underscores a crucial finding: the model's testing accuracy is contingent upon training with the same inclination. For instance, the model was trained with an inclination of 1:20 at a speed of 130 km/h and tested on a similar track and speed but with an inclination of 1:30. The figure vividly portrays the failure in detecting lateral irregularities, manifested by a substantial disparity between real and predicted values, indicative of notably low accuracy.

Hence, it was observed that the model exhibited optimal performance when trained and tested under consistent rail inclinations. The accuracy of the model proved to be dependent on maintaining a uniform inclination between the training and testing phases. Therefore, when the inclination remains constant during both phases, the model can effectively identify lateral irregularities even in the presence of varying track characteristics and speeds.

Consequently, a categorization procedure was started, dividing the data into three different groups according to different inclinations. Every group had distinct phases of training and testing, with consistent inclinations within the respective group. Using Random Forest Regression, one group's features best and worst performances. This algorithm was selected because it performed better at inclinations of 1:20, 1:30, and 1:40 than Polynomial Regression and Support Vector Regression. The subsequent sections of this thesis will predominantly present results based on Random Forest Regression owing to its heightened accuracy and efficiency.

It should be noted that similar special limitations have not been found for other variables than inclination, like different characteristics for different tracks. Therefore, it is possible to test the case with the model trained on another track with different characteristics and speed but the same inclination.

- La sommatoria  $\sum_{k=1}^k s_{k,j}$  indica la media delle metriche per le k permutazioni;
- k si riferisce al numero di permutazioni eseguite per la caratteristica. Essa conta il numero di volte che la caratteristica  $f_j$  viene permutata e la metrica  $s_{(k,j)}$  viene calcolata.

5. Risultati e discussione

5.1. Influenza dell'inclinazione della rotaia sulla stima delle irregolarità laterali

IDurante l'analisi dei dati di output della simulazione Gensys® per la creazione e lo sviluppo del modello, è emerso un dato significativo, che riguarda l'influenza dell'inclinazione della rotaia sul modello. La Fig. 4 illustra il tentativo di validazione iniziale, che ha portato a un fallimento. Questo risultato evidenzia un aspetto cruciale: l'accuratezza del test del modello è subordinata all'addestramento con la stessa inclinazione. Ad esempio, il modello è stato addestrato con un'inclinazione di 1:20 a una velocità di 130 km/h e testato su un tracciato e una velocità simili ma con un'inclinazione di 1:30. La figura rappresenta chiaramente l'insuccesso nel rilevamento delle irregolarità laterali, manifestato da una sostanziale disparità tra valori reali e previsti dal modello, indicativa di un'accuratezza molto bassa di quest'ultimo.

Il modello ha mostrato prestazioni ottimali quando addestrato e testato con inclinazioni coerenti. L'accuratezza del modello si è dimostrata dipendente dal mantenimento di un'inclinazione uniforme tra le fasi di addestramento e di test: quando l'inclinazione rimane costante durante entrambe le fasi, il modello può identificare efficacemente le irregolarità laterali, anche in presenza di caratteristiche e velocità variabili del binario. Di conseguenza, è stata avviata una procedura di categorizzazione, dividendo i dati in tre gruppi diversi in base alle diverse inclinazioni. Ogni gruppo ha avuto fasi distinte di addestramento e test, con inclinazioni coerenti all'interno del rispettivo gruppo. Utilizzando la regressione *Random Forest*, sono state individuate le caratteristiche migliori e peggiori di ciascun gruppo. Questo algoritmo è stato selezionato perché ha funzionato meglio con le inclinazioni di 1:20, 1:30 e 1:40 rispetto alla *Regressione Polinomiale* e alla *Support Vector Machine Regression*. Nel seguito dell'articolo si presenteranno principalmente risultati basati sulla regressione *Random Forest*, grazie alla sua maggiore accuratezza ed efficienza.

5.2. Data analysis and visualisations

The sample model, trained with an inclination of 1:20 and a speed of 130 km/h, was used to test the different case studies which could be seen in the Tab. 2. Among the case studies, the highest accuracy was achieved for the ST1 track at a speed of 160 km/h, while the ST1 track at 210 km/h had the worst performance in this group. Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9 and Fig. 10 show the accuracy results for the 160 km/h and 210 km/h case studies for the sections of whole track, straight and curve, respectively.

The data analysis presented in the figures above revealed the successful identification of lateral irregularity for group 1, which is the one with the inclination of 1:20; as it has the best accuracy for speed of 160 km/h, with R2 scores of 0.92, 0.90, and 0.94 for the entire track, straight track, and curve track, respectively. The accuracy levels were found to be high, but as the speed of the train increased, the accuracy of the curve began to drop due to greater lateral displacement. The accuracy of the entire track also decreased with higher speeds, while the accuracy of the straight track remained almost constant. Therefore, it is recommended to focus on overall track accuracy while considering the speed fluctuation.

Similar accuracy-speed trends were observed for the other groups with inclinations of 1:30 and 1:40, suggesting that the speed limit for the functionality of this study should be

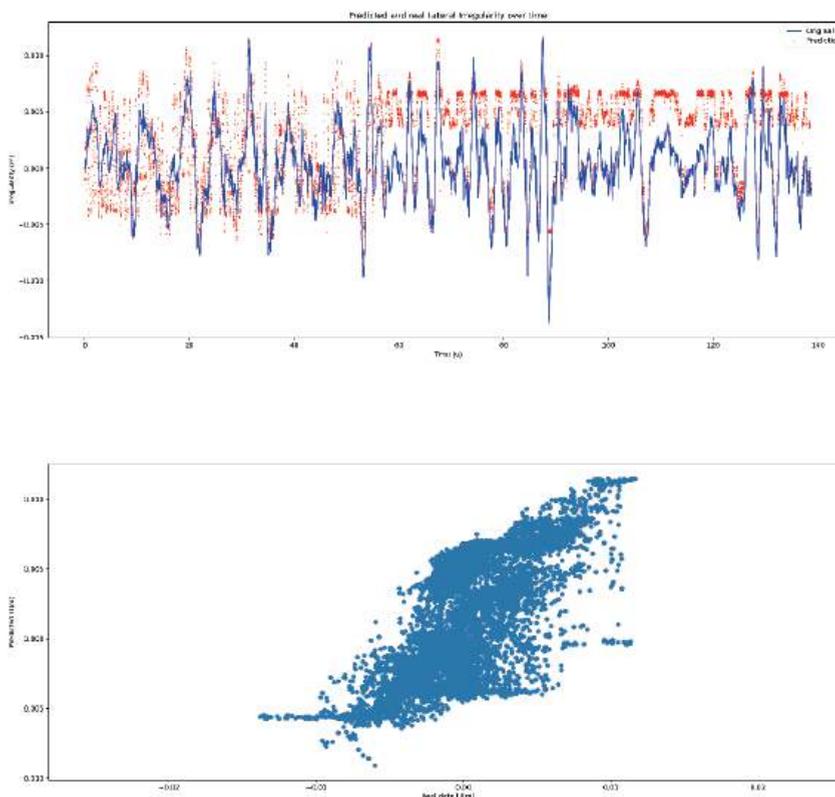


Figura 4 – Validazione fallita.  
Figure 4 – Failed validation.

Caratteristiche dei casi di studio  
Case Studies Characteristics.

| Caso di studio<br>Case Studies | Velocità<br>(km/h)<br>Speed (km/h) | Inclinazione<br>Inclination | Profilo della<br>ruota<br>Wheel Profile | Profilo del<br>binario<br>Rail Profile | Curve<br>Curves  | Lunghezza del<br>tracciato (m)<br>Track Length<br>(m) |
|--------------------------------|------------------------------------|-----------------------------|---|--|------------------|---|
| ST1                            | 130, 160, 210                      | 1:20, 1:30, 1:40            | EN S1002                                | 60 EI                                  | 2                | 13,276  |
| ST2                            | 50, 80, 120                        | 1:20, 1:30, 1:40            | EN S1002                                | 60 EI                                  | Molte<br>Several | 7,725   |

Va notato che un problema simile non è stato riscontrato per variabili diverse dall'inclinazione, come ad esempio le altre caratteristiche del binario. Pertanto, è possibile testare il modello addestrato su un binario con altre caratteristiche e percorso a velocità diverse ma che presenti la stessa inclinazione del binario di test.

5.2. Analisi e visualizzazione dei dati

Il modello campione, addestrato con un'inclinazione di 1:20 e una velocità di 130 km/h, è stato utilizzato per testare i diversi casi di studio descritti nella Tab. 2. Tra i casi di studio, la massima accuratezza è stata raggiunta per il tracciato ST1 a una velocità di 160 km/h, mentre il tracciato ST1 a 210 km/h ha avuto la peggiore prestazione. In Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10 sono mostrati i risultati di accuratezza per i casi di studio a 160 km/h e 210 km/h per i tre casi relativi all'intero tracciato, al tratto rettilineo e al tratto in curva.

L'analisi dei dati presentata nelle figure ha rivelato l'identificazione riuscita delle irregolarità laterali per il gruppo 1, che è quello con l'inclinazione di 1:20, poiché ha la migliore accuratezza per la velocità di 160 km/h, con punteggi R2 di 0,92, 0,90 e 0,94 per l'intero tracciato, il tratto rettilineo e quello in curva, rispettivamente. I livelli di accuratezza riscontrati sono stati elevati, ma all'aumentare della velocità del veicolo l'accuratezza del tratto in curva ha iniziato a diminuire a causa dei maggiori spostamenti laterali. Anche l'accuratezza dell'intero tracciato è diminuita a velocità più elevate, mentre l'accuratezza del tratto rettilineo è rimasta pressoché costante. Pertanto, si raccomanda di concentrarsi sull'accuratezza complessiva del tracciato tenendo conto della variazione della velocità. Simili andamenti dell'accuratezza in funzione della velocità sono stati osservati per gli altri gruppi con inclinazioni di 1:30 e 1:40, il che suggerisce che i risultati di questo

around 210 km/h. Among the tested models, the ST1 with a speed of 160 km/h exhibited the best accuracy with R2 of 0.92 (92% prediction accuracy), and an RMSE of 0.000930. The other tested models in this group of 1:20 inclinations, also showed good accuracy levels, except for the ST1 with a speed of 210 km/h, which had the worst scenario with R2 of 0.80 (80% accuracy) and an RMSE of 0.001514. The ST2 had a maximum accuracy of R2 = 0.91 at a speed of 50 km/h and a minimum accuracy of R2 = 0.78 at a speed of 120 km/h.

Analysing all the error distribution of the prediction on the histograms with the percentiles of 1 and 99, is illustrating that the 99% of the errors are in average distributed between ±2 mm, which is acceptable for our case study.

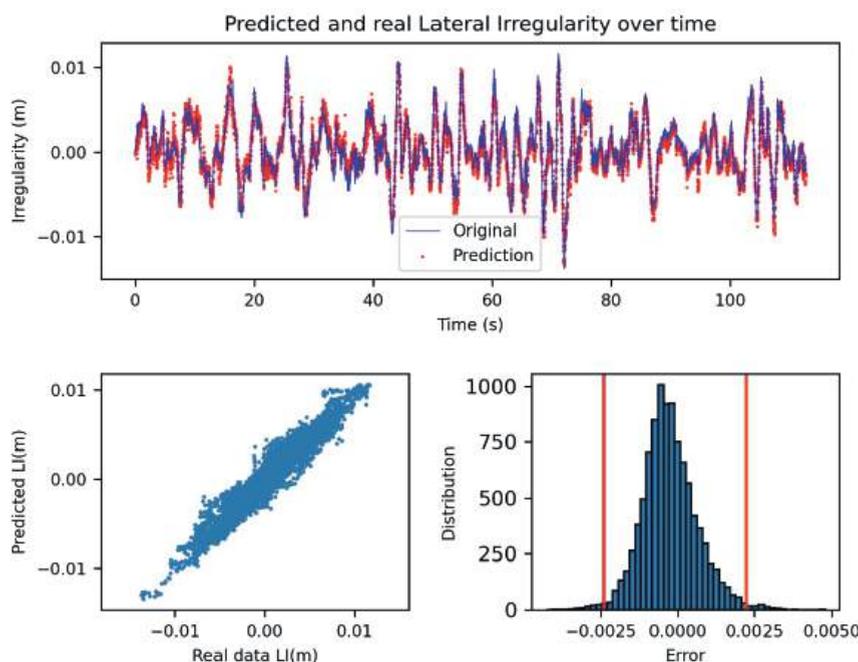


Figura 5 – Irregolarità laterali previste e reali nel tempo e distribuzione degli errori per l'intero tracciato testato sul percorso Velim (ST1) con inclinazione di 1:20 a 160 km/h.

Figure 5 – Predicted and real Lateral Irregularity over time, and error distribution for the whole track tested on the Velim Track (ST1) with the Inclination of 1:20, 160 km/h.

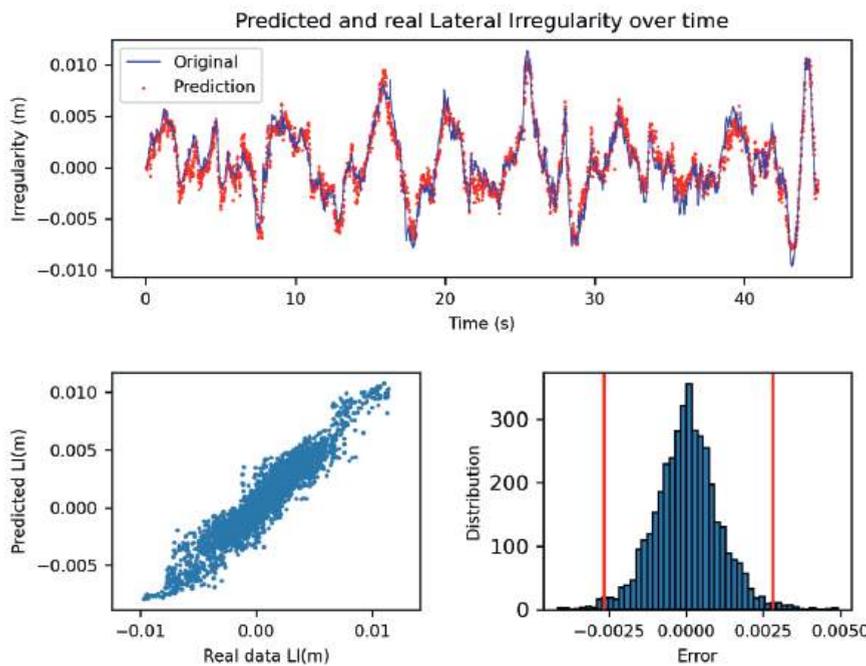


Figura 6 – Irregolarità laterali previste e reali nel tempo e distribuzione degli errori per il tratto rettilineo testato sul percorso Velim (ST1) con inclinazione di 1:20 a 160 km/h.

Figure 6 – Predicted and real Lateral Irregularity over time and error distribution for the Straight track tested on the Velim Track (ST1) with the Inclination of 1:20, 160 km/h.

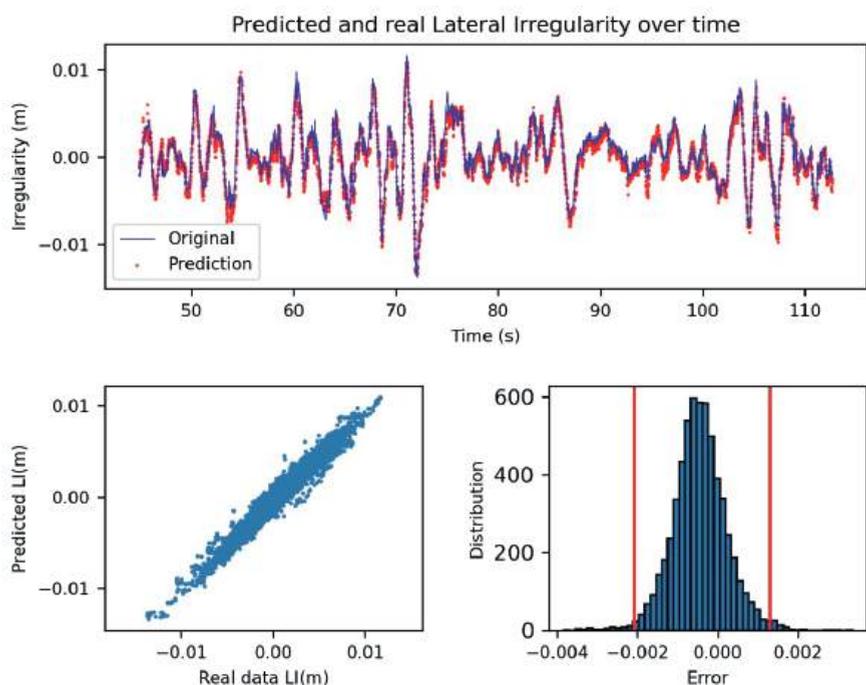


Figura 7 – Irregolarità laterali previste e reali nel tempo e distribuzione degli errori per il tratto in curva testato sul percorso Velim (ST1) con inclinazione di 1:20 a 160 km/h.

Figure 7 – Predicted and real Lateral Irregularity over time and error distribution for the Curve track tested on the Velim Track (ST1) with the Inclination of 1:20, 160 km/h.

To address concerns regarding the decrease in accuracy at higher speeds, potential convergence issue was considered and an investigation into convergence behaviour during the training phase was conducted, monitoring key metrics such as loss function values and validation accuracy. Training curves were examined for signs of instability or erratic fluctuations.

The selected algorithms and learning rates were explored to assess their impact on convergence. The persistent decrease in accuracy suggested that the challenge at higher speeds is not due to convergence issues but may be influenced by the inherent complexities of the problem domain. Hence, as the convergence issue was checked, and it did not have any effect on the accuracy, the results are reliable.

After aggregating all data and effective features, the main model was created and trained with a test size of 0.07, which means that 30% of the data was used for training while 70% was used for the testing. The results were satisfactory for both Random Forest and polynomial regressions. In Fig. 11, one example of the error distribution of each model is showed, and still the error is distributed with the same values as before for both different algorithms. The total average accuracy is 0.921 (R2 Score) for RFR and 0.926 for polynomial regressions.

### 5.3. Sensitivity Analysis Result

For sensitivity analysis, feature importance analysis was conducted to detect the relative significance and contribution of each input variable in the model's predictive performance. This analysis results for the random forest model reveal that the most significant impact on the model is attributed to lateral displacement, while the least crucial feature among the considered variables is associated with lateral acceleration. This outcome, though surprising, aligns with the initial notion that there is no discernible relationship between lateral acceleration and lateral irregularity in the track due to the variables between the wheel and the rail in the lateral direction. Following LDWR, curvature emerges as the next most influential parameter, with the remaining variables contribute to a lesser extent.

studio sono attendibili fino a una velocità massima di circa 210 km/h. Tra i modelli testati, l'ST1 con una velocità di 160 km/h ha mostrato la migliore accuratezza con R2 di 0,92 (accuratezza di previsione del 92%) e un RMSE di 0,000930. Anche gli altri modelli testati in questo gruppo di 1:20 hanno mostrato buoni livelli di accuratezza, eccetto il ST1 con una velocità di 210 km/h, che ha avuto lo scenario peggiore con R2 di 0,80 (accuratezza dell'80%) e RMSE di 0,001514. Il ST2 ha avuto un'accuratezza massima di R2 = 0,91 a una velocità di 50 km/h e un'accuratezza minima di R2 = 0,78 a una velocità di 120 km/h.

Analizzando la distribuzione degli errori della previsione sugli istogrammi con i percentili di 1 e 99, si dimostra che il 99% degli errori è in media distribuito tra  $\pm 2$  mm, il che è accettabile per il nostro caso di studio. Per affrontare le preoccupazioni relative alla diminuzione dell'accuratezza a velocità più elevate, è stato preso in considerazione un potenziale problema di convergenza ed è stata condotta un'indagine sul comportamento della convergenza durante la fase di addestramento del modello, monitorando parametri chiave come i valori della funzione di perdita e la precisione di convalida. Le curve di addestramento sono state esaminate per individuare segnali di instabilità o fluttuazioni irregolari.

Gli algoritmi selezionati e i tassi di apprendimento sono stati indagati per valutare il loro impatto sulla convergenza. Il persistere della diminuzione di accuratezza ha suggerito che le problematiche a velocità più elevate non siano dovute a problemi di convergenza, ma possano essere influenzate dalle complessità intrinseche del dominio del problema. Quindi, poiché la questione della convergenza è stata controllata dimostrando di non avere alcun effetto sull'accuratezza, i risultati si sono ritenuti affidabili.

Dopo aver aggregato tutti i dati e le caratteristiche efficaci, il modello principale è stato creato e addestrato con una dimensione di test di 0,07, il che significa che il 30% dei dati è stato utilizzato per l'addestramento mentre il 70% è stato utilizzato per il test. I risultati sono stati soddisfacenti sia per *Random Forest Regression* (RFR), sia per la *Regression*

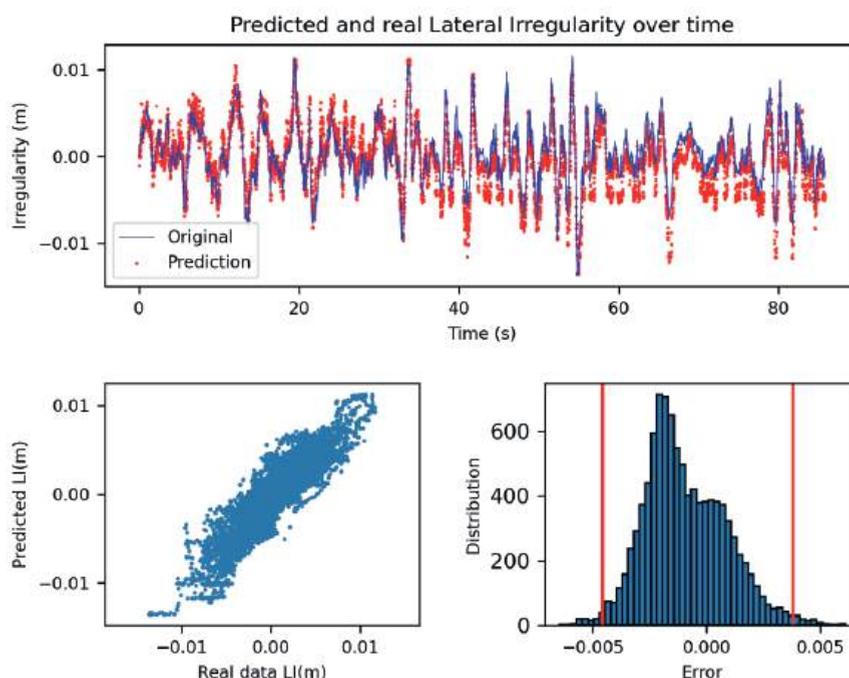


Figura 8 – Irregolarità laterali previste e reali nel tempo e distribuzione degli errori per l'intero tracciato testato sul percorso Velim (ST1) con inclinazione di 1:20 a 210 km/h.

Figure 8 – Predicted and real Lateral Irregularity over time and error distribution for the whole track tested on the Velim Track (ST1) with the Inclination of 1:20, 210 km/h.

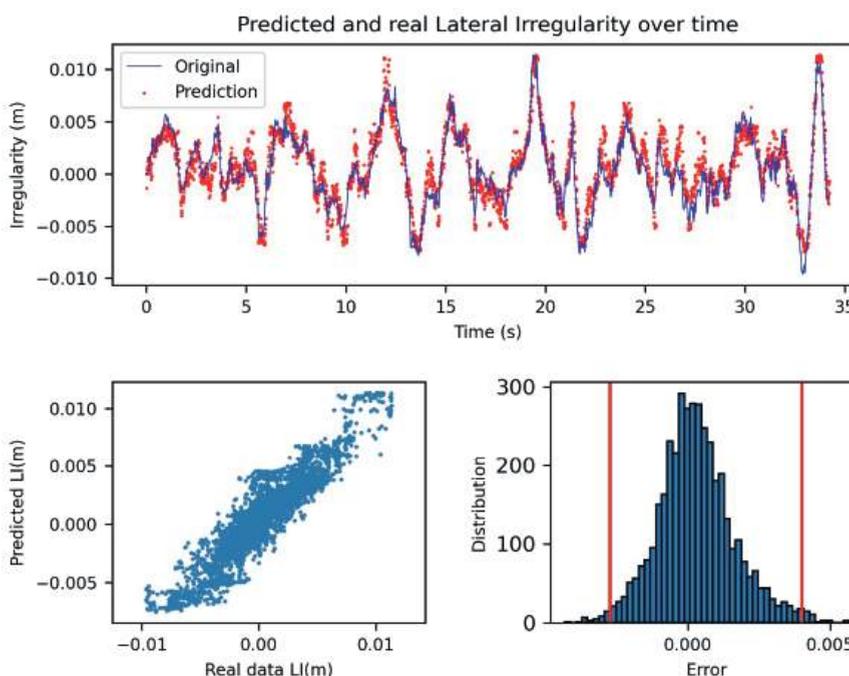


Figura 9 – Irregolarità laterali previste e reali nel tempo e distribuzione degli errori per il tratto rettilineo testato sul percorso Velim (ST1) con inclinazione di 1:20 a 210 km/h.

Figure 9 – Predicted and real Lateral Irregularity over time and error distribution for the Straight track tested on the Velim Track (ST1) with the Inclination of 1:20, 210 km/h.

Polimoniale. Nella Fig. 11, si può vedere un esempio della distribuzione dell'errore di ogni modello, ancora distribuito con gli stessi valori per entrambi gli algoritmi. L'accuratezza media totale è 0,921 (punteggio R2) per RFR e 0,926 per la regressione polinomiale.

**5.3. Risultati dell'analisi di sensibilità**

Riguardo all'analisi di sensibilità, è stata condotta un'analisi dell'importanza delle caratteristiche per comprendere l'importanza relativa e il contributo di ciascuna variabile di input nelle prestazioni predittive del modello. I risultati di questa analisi per il modello *Random Forest* hanno mostrato che l'influenza più significativa sul modello è degli spostamenti laterali, mentre la caratteristica meno significativa tra le variabili considerate è l'accelerazione laterale. Questo risultato, sebbene sorprendente, è in linea con l'osservazione iniziale secondo cui non esiste una relazione discernibile tra accelerazioni e irregolarità laterali del binario a causa delle diverse variabili che influenzano il moto relativo della ruota rispetto alla rotaia nella direzione laterale. Dopo gli spostamenti laterali, la curvatura del tracciato emerge come il parametro più influente, con le variabili rimanenti che contribuiscono in misura minore. È significativo notare che i test numerici sono stati condotti sia tenendo conto della conicità della ruota sia non considerandola. In assenza di conicità, l'accelerazione laterale assume un ruolo più importante, fornendo il contributo che altrimenti sarebbe fornito dalla conicità. Sorprendentemente, anche senza considerare la conicità, il rilievo delle irregolarità rimane accettabile. Considerare la conicità ha migliorato le prestazioni della *Random Forest Regression* (RFR), mentre il suo effetto sulla regressione polinomiale è risultato relativamente modesto. La Fig. 12 fornisce una rappresentazione grafica di questi risultati.

**5.4. Analisi di accuratezza considerando l'inclinazione della rotaia, la velocità e l'effetto dell'algoritmo**

Il confronto delle accuratèzze ottenute con la *Random Forest Regression*

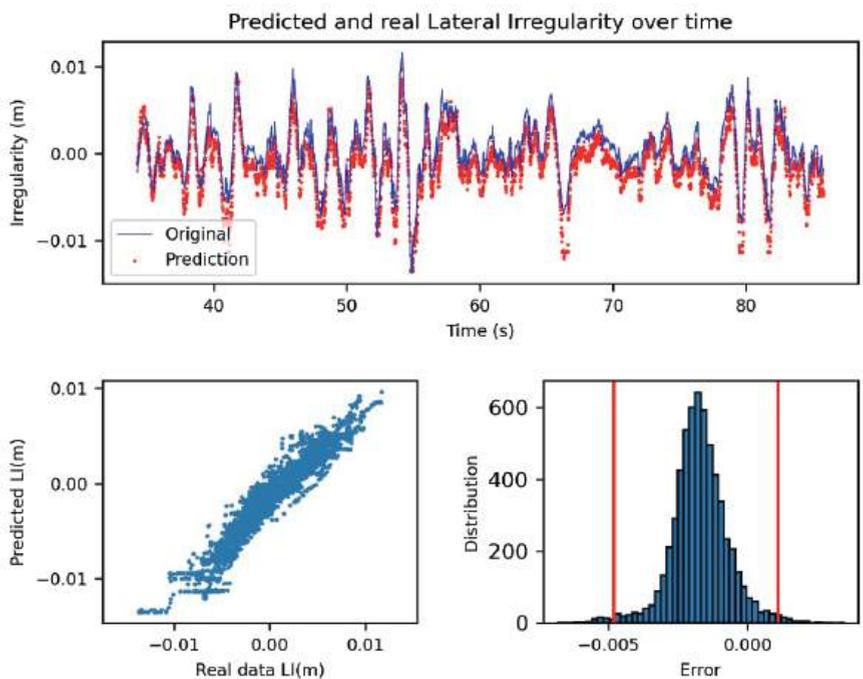


Figura 10 – Irregolarità laterali previste e reali nel tempo e distribuzione degli errori per il tratto curvo testata sul percorso Velim (ST1) con inclinazione di 1:20 a 210 km/h.

Figure 10 – Predicted and real Lateral Irregularity over time and error distribution for the Curve track tested on the Velim Track with the Inclination of 1:20, 210 km/h.

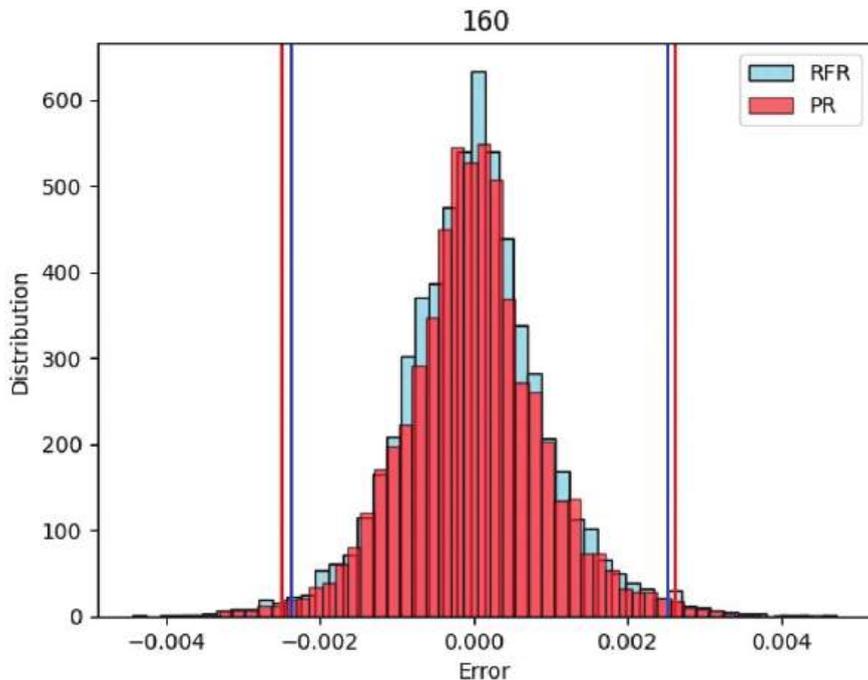


Figura 11 – Distribuzione degli errori per la velocità di 160 km/h testata con RFR e regressione polinomiale.

Figure 11 – Error Distribution for the speed of 160 km/h tested RFR and PR.

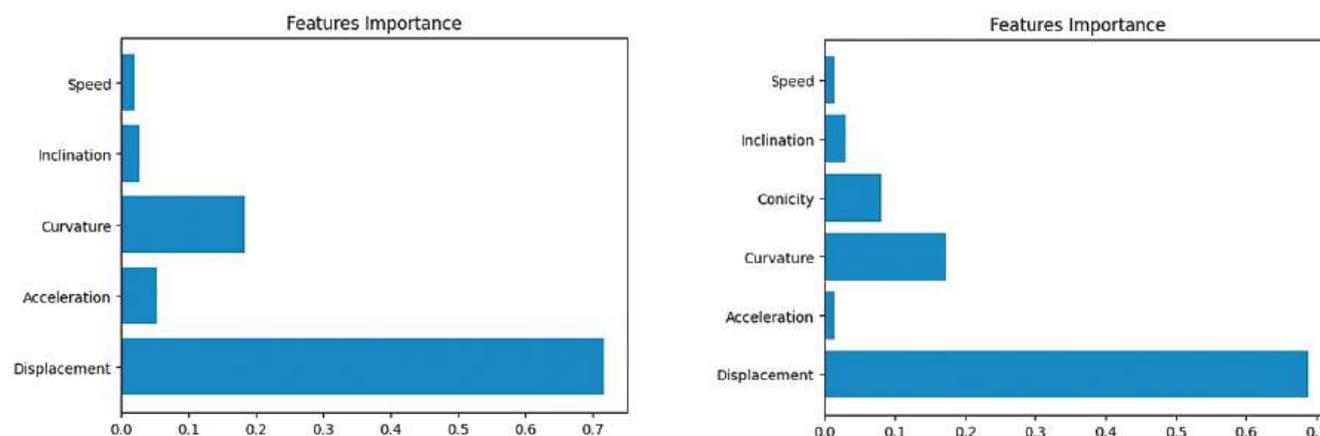


Figura 12 – Analisi delle caratteristiche del modello *Random Forest* senza conicità (sinistra), con conicità (destra).  
 Figure 12 – Feature analysis of *Random Forest* model without Conicity (Left) and with Conicity (Right).

(RFR) e con la regressione polinomiale è stato condotto in funzione della velocità e all'inclinazione della rotaia, come illustrato nelle Fig. 13 e Fig. 14. I risultati rivelano un andamento in cui all'aumento della velocità corrisponde una diminuzione dell'accuratezza. Inoltre, anche l'inclinazione della rotaia influenza l'accuratezza, con inclinazioni maggiori che contribuiscono a una maggiore stabilità nello spostamento delle ruote sulla rotaia. In particolare, l'inclinazione di 1:40 mostra gli intervalli di accuratezza più elevati, seguita da 1:30, mentre 1:20 mostra l'accuratezza più bassa. Sebbene l'accuratezza diminuisca a velocità più elevate per tutte le inclinazioni, con inclinazioni di 1:40 e 1:30 rimane comunque accettabile, superando l'80%, mentre l'intervallo di accuratezza per l'inclinazione di 1:20 scende al di sotto del 75%, suggerendo l'opportunità di ulteriori indagini in ricerche future. Il punteggio R2 medio complessivo per entrambi gli algoritmi è di circa 0,88 in

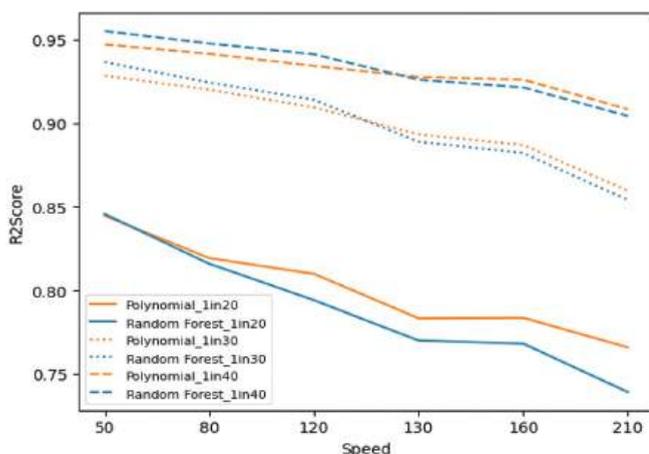


Figura 13 – Accuratezza in funzione di velocità e inclinazione senza conicità.  
 Figure 13 – Accuracies based on Speed and Inclination without Conicity.

*It is significant to note that the numerical tests conducted both with and without consideration of conicity. Interestingly, in the absence of conicity, acceleration assumes a more prominent role, taking the contribution that would otherwise be attributed to conicity. Strikingly, even without measuring conicity, the detection of irregularities remains feasible. The impact of conicity is observed as an enhancement in the performance of random forest regression (RFR), while its effect on polynomial regression (PR) is comparatively minimal. Fig. 12 provide with visual representations of these findings.*

#### 5.4. Accuracy analysis Considering Rail Inclination, Speed and Algorithmic Impact

*The comparison of accuracies between Random Forest Regression (RFR) and polynomial regression was conducted based on speed and rail inclination, as illustrated in Fig. 13 and Fig. 14. The findings reveal a trend where increasing speed corresponds to a decrease in accuracy. Moreover, the rail inclination also influences accuracy, with higher inclinations contributing to greater stability in wheel displacement on the rail.*

*Specifically, the inclination of 1:40 exhibits the highest accuracy ranges, followed by 1:30, while 1:20 demonstrates the lowest accuracy. Although the accuracy drops in higher speed for all inclinations but at inclinations of 1:40 and 1:30 remains acceptable, exceeding 80%, and the accuracy range for the inclination of 1:20 falls below 75%, warranting further investigation in future research. The overall average R2 Score for both algorithms is approximately 0.88 in the absence of conicity. In the presence of conicity, the R2 Score is 0.96 for RFR and 0.86 for PR.*

*Since most of the results of this work was focused on the Random Forest Regression algorithm, in Fig. 15 accuracies based on the rail inclination and speed considering conicity specifically for illustrating, and comparing the Random Forest with other potential algorithms.*

assenza di conicità. In presenza di conicità, il punteggio R2 è 0,96 per la RFR e 0,86 per regressione polinomiale.

Poiché la maggior parte dei risultati di questo lavoro si è concentrata sull’algoritmo *Random Forest Regression*, nella Fig. 15 l’andamento delle accuratze in funzione della velocità e dell’inclinazione della rotaia, che considera la conicità, è riportato per la *Random Forest Regression* e confrontato con gli altri algoritmi.

6. Conclusioni

Questo studio ha confermato la possibilità di rilevare le irregolarità laterali del binario tramite gli spostamenti laterali della ruota rispetto alla rotaia grazie alla loro correlazione. Ciò potrebbe aprire opportunità per lo sviluppo di sistemi di monitoraggio a bordo, privi di piattaforma inerziale, e adatti ad essere installati a bordo dei treni in servizio commerciale. L’importante volume di dati raccogliabile quotidianamente dai vari treni commerciali offrirebbe la possibilità di sviluppare le pratiche di manutenzione predittiva. Dopo aver testato diversi algoritmi di apprendimento supervisionato, i risultati hanno mostrato che la regressione *Random Forest* e la Regressione Polinomiale hanno entrambe un’eccellente accuratezza, dimostrando la loro idoneità per il rilevamento delle irregolarità laterali in base agli spostamenti laterali della ruota. La *Support Vector Machine Regression*, d’altro canto, si è dimostrata inadeguata per questo scopo. Le caratteristiche essenziali per la costruzione del modello sono state classificate nel seguente ordine: spostamenti laterali, curvatura, conicità (se applicabile), accelerazione laterale e velocità. Questi risulti sono significativi, considerando che gran parte della ricerca precedente si è concentrata principalmente sulla ricostruzione delle irregolarità laterali del binario sulla base delle accelerazioni. È fondamentale sottolineare che l’inclinazione di posa della rotaia ha svolto un ruolo significativo nelle simulazioni e nello sviluppo degli algoritmi; pertanto, il suo valore non deve variare per le fasi di ad-

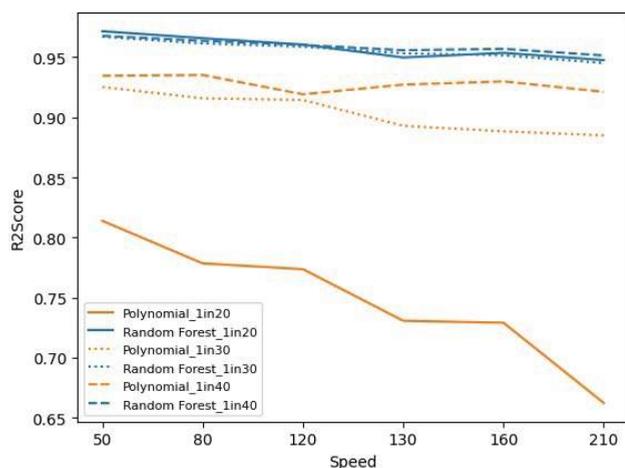


Figura 15 – Accuratezze in funzione di velocità e inclinazione con conicità per l’algoritmo *Random Forest Regression*.  
 Figure 15 – Accuracies based on Speed and Inclination with conicity for *Random Forest Regression* Algorithm.

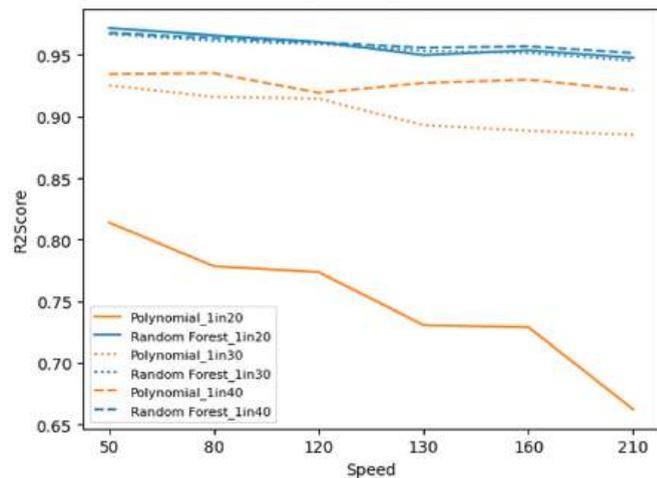
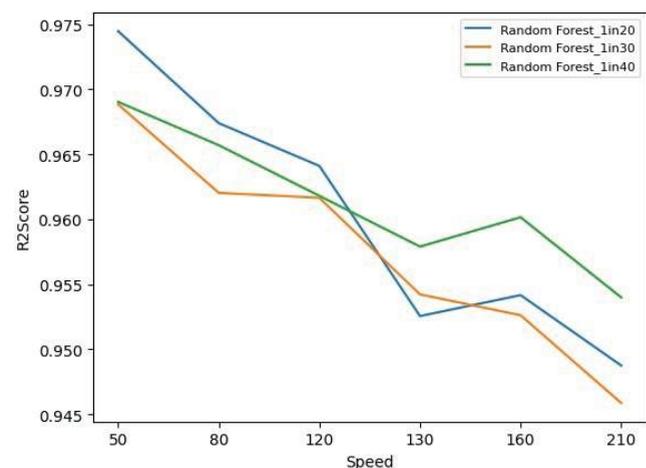


Figura 14 – Accuratezza in funzione di velocità e inclinazione con conicità.  
 Figure 14 – Accuracies based on Speed and Inclination with Conicity.

6. Conclusion

*This study has demonstrated the feasibility of detecting lateral irregularities through the lateral displacement of the Wheel/Rail with their correlation. This opens opportunities for the development of an on-board monitoring system, eliminating the need for inertial platforms and enabling frequent monitoring. The substantial volume of data collected daily from various commercial trains holds the potential for advancing predictive maintenance practices.*

*After testing several supervised learning algorithms, the results showed that random forest regression and polynomial regression offer both excellent accuracies, proving their suitability for lateral irregularity detection based on the wheel’s lateral displacement. Support vector machines, on the other hand, proved inadequate for this purpose.*



destramento e di test del modello. Mentre l'inclinazione della rotaia potrebbe inizialmente non sembrare un fattore significativo nelle valutazioni della dinamica di marcia, questo studio rivela il suo impatto sul comportamento in condizioni normali di marcia, come evidenziato dai modelli *Random Forest* e di regressione polinomiale.

La riduzione dell'accuratezza della previsione per l'allineamento all'aumentare dell'inclinazione della rotaia, specialmente nel caso di un'inclinazione di 1:20, è attribuita ad ampiezze di oscillazione laterale causate da più importanti movimenti di *body hunting* a bassa frequenza. La conicità equivalente eccezionalmente bassa della ruota S1002, abbinata all'inclinazione della rotaia di 1:20, fa sì che questi movimenti diventino potenzialmente instabili, a meno che i parametri del modello non siano attentamente selezionati.

Si prevede che questi risultati possano supportare il miglioramento delle ipotesi di modellazione quando si utilizzano la simulazione multi-corpo (MBS) per la generazione di set di dati di addestramento per monitoraggio. Si sottolinea inoltre l'importanza di utilizzare set di dati misurati per l'addestramento ML e di limitare la simulazione alle attività preliminari di *screening*.

L'impatto della velocità sull'accuratezza è risultato evidente nell'analisi delle sezioni in curva, con l'aumento della velocità che ha comportato una riduzione dell'accuratezza per entrambi gli algoritmi e per tutte le inclinazioni della rotaia. L'algoritmo sviluppato in questo studio ha dimostrato affidabilità fino a una velocità di 210 km/h.

I risultati di questo lavoro potrebbero rappresentare un passo avanti verso la manutenzione predittiva del binario. Si è dimostrato, infatti, che risulta possibile rilevare le irregolarità laterali del binario in funzione degli spostamenti laterali relativi ruota-rotaia. Pertanto, risolvendo il problema della misura da bordo treno delle irregolarità laterali, tenendo conto dell'effetto delle accelerazioni, è possibile ottenere un sistema completo per il rilievo della geometria del binario. Quindi, monitorando lo stesso binario più volte al giorno da treni in servizio commerciale, sarà possibile ottenere una grande quantità di dati utili a studiare l'evoluzione temporale delle caratteristiche geometriche, rendendo possibile un monitoraggio predittivo affidabile.

*The crucial features essential for constructing the model were ranked in the following order: Lateral displacement, curvature, conicity (if applicable), acceleration, and speed. This finding is noteworthy, considering that much of the previous research has primarily concentrated on reconstructing irregularities based on acceleration. It is imperative to highlight that the inclination played a significant role in both simulations and the development of algorithms, and its value must remain consistent for both the training and testing phases. While rail inclination might not initially appear to be a significant factor in running dynamics assessments, this study reveals its impact on the behaviour under normal running conditions, as identified by the Random Forest and polynomial regression models.*

*The reduction in prediction accuracy for lateral alignment with higher rail inclination, especially in the case of a 1:20 inclination, is attributed to amplified lateral oscillation amplitudes caused by low-frequency body-hunting motions. The exceptionally low equivalent conicity of the S1002 wheel, coupled with the 1:20 rail inclination, results in these motions potentially becoming unstable unless the model parameters are carefully selected.*

*This finding is expected to support improved modelling assumptions when using MBS for the generation of training datasets for condition monitoring developments. It also stresses the importance of using measured datasets for definitive ML training and limiting simulation to preliminary screening activities.*

*The impact of speed on accuracy was evident in the analysis of curved sections, with higher speeds resulting in decreased accuracy for both algorithms and across all inclinations. The algorithm developed in this study demonstrated reliability up to a speed of 210 km/h.*

*Finally, the result of this work could be a step toward the predictive maintenance. It was illustrated that it is possible to detect the lateral irregularities based on displacements and by solving this challenge of monitoring the lateral irregularity through on-board measurement by adding the effect of acceleration, it is possible to have a full track geometry system. Hence, by monitoring the same track every day, it will be possible to get a large amount of data, which gives the chance of studying the temporal evolution of geometrical characteristics and the predictive maintenance would be the new challenge for the future work.*

#### BIBLIOGRAFIA - REFERENCES

- [1] SOLEIMANMEIGOUNI I. (2019), "Predictive Models for Railway Track Geometry Degradation" – Doctoral Thesis, Luleå University of Technology.
- [2] CÁRDENAS-GALLO I., SARMIENTO C.A., MORALES G.A., BOLIVAR M.A., AKHAVAN TABATABAEI R. (2017), "An ensemble classifier to predict track geometry degradation. Reliability Engineering & System Safety" - Reliability Engineering and System Safety, 161(C), 53-60 (DOI: 10.1016/j.ress.2016.12.012).
- [3] ROSANO G., MASSINI D., BOCCIOLINI L., ZAPPACOSTA C., DI GIALLEONARDO E., SOMASCHINI C., LA PAGLIA I., PUGI L. (2024), "Diagnostics of the railway track - Possibility of development through the measurement of accelerations and contact forces" - Ingegneria Ferroviaria, 2 (DOI: 10.57597/IF.02.2024.ART.1).

- [4] European Committee for standardization (2019), “EN 13848-1: Railway applications - Track - Track geometry quality - Part 1: Characterization of track geometry”.
- [5] MANCINI M. (2007), “Aspetti della qualità geometrica del binario”, *Ingegneria Ferroviaria*, 2”.
- [6] IN2SMART Project (2019), “Report on track/switch parameters and problem zones” - Technical report, Deliverable D4.1 (DOI: 10.3030/730569).
- [7] WESTON P., LING C.S., GOODMAN C.J., ROBERTS C., LI P., GOODALL R.M. (2007), “Monitoring lateral track irregularity from in-service railway vehicles - Proceedings of the Institution of Mechanical Engineers”, Part F: Journal of Rail and Rapid Transit, 221(1), 89-100 (DOI: 10.1243/0954409JRRT64).
- [8] WESTON P., ROBERTS C., YEO G., STEWART E. (2015), “Perspectives on railway track geometry condition monitoring from in-service railway vehicles” - *Vehicle System Dynamics* 53(7), 1063-1091 (DOI: 10.1080/00423114.2015.1034730).
- [9] ASSETS4RAIL Project (2020), “Wheel-rail transversal position monitoring technologies” - Technical Report, Deliverable D8.1 (<http://www.assets4rail.eu/results-publications/>).
- [10] SHI D., SABANOVIC E., RIZZETTO L., SKRICKIJ V., OLIVERIO R., KAVIANI N., YE Y., BUREIKA G., RICCI S., HECHT M. (2022), “Deep learning based virtual point tracking for real-time target-less dynamic displacement measurement in railway applications” - *Mechanical Systems and Signal Processing*, 166, 108482, 2022 (doi: 10.1016/j.ymssp.2021.108482).
- [11] TRUE H., CHRISTIANSEN L.E., PLESNER A., LONSTRUP AMMITZBOLL A., JERRAM DAHL B. (2017), “Why is it so difficult to determine the lateral Position of the Rails by a Measurement of the Motion of an Axle on a moving Vehicle? – Proceedings of First International Conference on Rail Transportation”, Chengdu (DOI: 10.1061/9780784481257.fm).
- [12] ROSA A.D., ALFI S., BRUNI S. (2019), “Estimation of lateral and cross alignment in a railway track based on vehicle dynamics measurements” - *Mechanical Systems and Signal Processing*. 116, 606-623 (DOI: 10.1016/j.ymssp.2018.06.041).
- [13] ROSA A.D., KULKARNI R., QAZIZADEH A., BERG M., DI GIALLEONARDO E., FACCHINETTI A., BRUNI S. (2020), “Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms”, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 235, 1 (DOI: 10.1177/0954409720906649).
- [14] ANTOGNOLI M., BUREIKA G., KAVIANI N., RICCI S., RIZZETTO L., SKRICKIJ V. (2021), “Development of a contactless sensor system to support rail track geometry on-board monitoring”, *Proceedings of CETRA 6th International Conference on Road and Rail Infrastructure*, Zagreb (DOI: 10.5592/CO/cetra.2020.1105).
- [15] ANTOGNOLI M., MARINACCI C., RICCI S., RIZZETTO L. (2020), “Requirement specifications for track measuring and monitoring systems” - *Ingegneria Ferroviaria*, 75, 11, 841-864.
- [16] MANNARA G. (2023) - “What can be done to make a measurement more and more useful?” (2023) - *Ingegneria Ferroviaria*, 78, 10 (DOI: 10.57597/IF.10.2023.ART.2)
- [17] GENSYS (2022), “Railroad applications”, ([https://www.gensys.se/doc\\_html/analyse\\_sum.html#Mainmenu](https://www.gensys.se/doc_html/analyse_sum.html#Mainmenu)).
- [18] VUZ (2022), “VUZ Test Centre Velim” (<https://www.cdvuz.cz/en/page/zkusebni-centrum-vuz-velim>).
- [19] IK (2022), “IK Test Centre Zmigrod” (<http://www.ikolej.pl/en/units/test-track-centre-near-zmigrod>).
- [20] UIC (2009), “Testing and approval of railway vehicles from the point of view of their dynamic behaviour - safety - track fatigue - ride quality”, UIC518: 4ED 2009.
- [21] EN (2205), “Railway applications - testing for the acceptance of running characteristics of railway vehicles - testing of running behaviour”, EN 14363:2005.
- [22] KOTSIANTIS S.B. (2007), “Supervised machine learning: a review of classification techniques” - *Informatica*, 31, 3-24.
- [23] LEE J.S., HWANG S.H., CHOI I.Y., KIM I.K. (2018), “Supervised machine learning: a review of classification techniques” - *Informatica*, 31, 3-24.
- [24] LUGARÀ A. (2018), “The railway predictive maintenance and the enabling role of the “Internet of Things””, *Ingegneria Ferroviaria*, 5, 434.
- [25] LYASHENKO V. (2022), “How to use Random Forest for regression: notebook, examples and documentation” – Intel Tiber AI Studio (<https://cnvrg.io/random-forest-regression/>).
- [26] BOBBITT Z. (2021), “RMSE vs. R-Squared: Which Metric Should You Use?” <https://www.statology.org/rmse-vs-rsquared>.

- [27] ZHAN D., JING D., WU M., ZHANG D., YU L., CHEN T. (2018), "An Accurate and Efficient Vision Measurement Approach for Railway Catenary Geometry Parameters" - IEEE Transactions on Instrumentation and Measurement, 67(12), 2841-2853 (DOI: 10.1109/TIM.2018.2830862).
- [28] TAO Z., HUILING L., WENWEN W., XIA Y. (2019), "GA-SVM based feature selection and parameter optimization in hospitalization expense modeling" - Applied Soft Computing, 75, 323-332 (DOI: 10.1016/j.asoc.2018.11.001).
- [29] ZHANG P. (2019), "A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model" - Applied Soft Computing, 85, 105859 (DOI:10.1016/j.asoc.2019.105859).
- [30] CAI J., LUO J., WANG S., YANG S. (2018), "Feature selection in machine learning: A new perspective". Neurocomputing, 300, 70-79 (DOI: 10.1016/j.neucom.2017.11.077).
- [31] PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M., PERROT M., DUCHESNAY E. (2011), "Scikit-learn: Machine learning in python" - Journal of Machine Learning Research, 12, 2825-2830 (DOI: 10.5555/1953048.2078195).

## La qualità è il nostro viaggio quotidiano.

### ISOTRACK

la divisione trasporti di ISOIL Industria Spa offre soluzioni di qualità da oltre vent'anni in ambito ferroviario.

#### LE NOSTRE APPLICAZIONI

- Treni ad Alta Velocità, Regionali e Mezzi d'Opera;
- Trasporto urbano su rotaia (metropolitane e tram) e gomma (autobus);
- Sistemi di sicurezza a bordo dei veicoli fino a SIL4.

#### SERVIZI DI QUALITÀ

- Riparazione qualificata della nostra strumentazione;
- Aggiornamento firmware display;
- Qualifica prodotti per sistemi Ready-to-Use;
- Supporto e consulenze per applicazioni specifiche.



[www.isoil.it](http://www.isoil.it)

ISOIL Industria s.p.a.  
Cinisello Balsamo - Milano (Italy)  
Via Fratelli Gracchi, 27  
tel. +39 0266027.1 - fax +39 026123202  
vendite@isoil.it

**ISOIL**  
INDUSTRIA  
Le soluzioni che contano



Costruzioni  
Linee  
Ferroviarie



Promofer  
Safety Services



**dal 1945**  
il futuro viaggia  
su binari sicuri



Strukton  
Rail