ISSN: 2519-9439

Data-Driven In-Train Force Monitoring Under Changing Consist Configurations and Loads

Ka-Yan Kan[1], Hiu-Tung Lam[2,] Hoi-Yan [Li1]

[1] Department of Electrical and Electronic Engineering, Faculty of Engineering, Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China

[2] Department of Electronic Engineering and Computer Science, School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, Hong Kong SAR, China

Abstract

The reliable monitoring of in-train forces is a critical component for ensuring the safety and efficiency of heavy-haul railway operations. Excessive longitudinal forces can lead to severe operational failures, including coupler breakage, train separation, and catastrophic derailments. Traditionally, these forces have been estimated using computationally expensive physics-based longitudinal train dynamics models, which struggle to adapt to real-time operations due to their reliance on static parameters and deterministic assumptions. This paper presents a comprehensive data-driven approach to monitor and predict in-train forces under dynamically changing consist configurations and varying payload distributions. By leveraging advanced machine learning algorithms, the proposed framework integrates multidimensional operational data, including locomotive control inputs, track topography, and historical sensor readings. We explore the transition from classical physical modeling to modern artificial intelligence techniques, evaluating models ranging from ensemble tree-based methods to recurrent neural networks designed for sequential time-series forecasting. Special attention is given to the challenge of varying train configurations, where the number, type, and weight of wagons change frequently, rendering static models obsolete. Through extensive analysis of simulated and real-world operational datasets, this research demonstrates that data-driven methodologies can achieve high accuracy in predicting draft gear forces and slack action occurrences. The findings provide a scalable, real-time solution for intelligent train control systems, ultimately contributing to enhanced railway safety, reduced maintenance costs, and optimized network capacity.

Keywords: In-train force, data-driven modeling, consist configuration, railway dynamics..

1.Introduction

1.1 Background and Motivation

The global demand for freight transportation has driven the railway industry to continually increase the length and mass of heavy-haul trains. As these trains become longer and heavier, the complexity of managing their longitudinal dynamics increases exponentially. In-train forces, which primarily consist of draft and buff forces transmitted through the couplers and draft gears connecting adjacent wagons, represent one of the most significant challenges in

ISSN: 2519-9439

modern railway engineering. When a train undergoes acceleration, braking, or traverses varying track gradients, the individual wagons experience different forces, leading to relative motion between them. This relative motion is accommodated by the draft gear, which acts as a shock absorber. However, when the kinetic energy exceeds the capacity of the draft gear, the resulting impact forces can be devastating. Severe in-train forces are the primary cause of coupler failure, where the mechanical connection between wagons breaks, leading to train separation. Even more critically, large compressive buff forces can push wagons off the track, resulting in catastrophic derailments. The economic and safety implications of such incidents are profound, necessitating rigorous monitoring and control of in-train forces as highlighted in early transportation studies. Historically, railway operators have relied on conservative driving strategies and stringent operational rules to mitigate the risk of excessive forces. Drivers are trained to apply traction and braking in a controlled manner, anticipating the response of the train based on route knowledge and experience. However, human intuition is limited, especially when dealing with trains exceeding several kilometers in length and tens of thousands of tons in mass. The physical characteristics of such trains are highly non-linear and distributed over a vast spatial domain. Therefore, the implementation of automated, intelligent monitoring systems is not just a technological luxury but an operational necessity. Such systems require continuous, accurate estimation of the forces occurring throughout the train to provide real-time feedback to the driver or the automated train operation system, as suggested by foundational research in train dynamics.

1.2 The Challenge of Changing Consist Configurations

A major complicating factor in the monitoring of in-train forces is the dynamic nature of train consist configurations. Unlike passenger trains, which often operate with fixed formations, freight trains are frequently reconfigured at shunting yards. A single train consist may be composed of a heterogeneous mix of wagon types, each with different aerodynamic profiles, braking systems, and draft gear characteristics. Furthermore, the loading condition of each wagon can vary dramatically from completely empty to fully loaded. The arrangement of these wagons, alongside the placement of distributed power locomotives within the consist, fundamentally alters the longitudinal dynamics of the entire train. A driving strategy that is perfectly safe for a uniformly loaded train may induce massive run-in forces if applied to a train with empty wagons positioned at the front and heavily loaded wagons at the rear. The constant variability in consist configurations makes it nearly impossible to rely on a single, static model for force prediction, a problem widely recognized in railway engineering literature. Traditional physics-based longitudinal train dynamics models attempt to solve this problem by mathematically simulating the entire train. These models require exact input parameters for every single component of the train, including the mass of each wagon, the non-linear stiffness and damping coefficients of every draft gear, the propagation delay of the pneumatic brake pipe, and the instantaneous friction coefficients at the wheel-rail interface. Obtaining these parameters with high accuracy in a real-world operational setting is practically impossible. Furthermore, even if the parameters were known, solving the vast array of coupled differential equations required for a multi-kilometer train in real-time demands substantial computational resources that are rarely available in on-board locomotive computing systems. To overcome these limitations, the railway industry is increasingly looking towards data-driven methodologies. By utilizing the massive amounts of data generated by modern locomotives equipped with advanced telemetry, global positioning systems, and on-board sensors, it is possible to train machine learning models that can infer the complex non-linear relationships between driver inputs, track conditions, and resulting in-train forces. These models can inherently learn the generalized dynamic behaviors associated with different consist configurations without requiring explicit mathematical

ISSN: 2519-9439

definitions of every physical component. This paper aims to thoroughly explore and validate the application of data-driven techniques for in-train force monitoring, explicitly addressing the challenges posed by continuously changing train consists and heterogeneous loading conditions.

2. Literature Review and Background

2.1 Evolution of Longitudinal Train Dynamics Modeling

The study of longitudinal train dynamics has a rich history, evolving from simple rigid-body calculations to highly complex, multi-body simulation frameworks. Early models treated the entire train as a single concentrated mass, which was sufficient for calculating overall energy consumption and basic performance metrics but completely inadequate for understanding the forces between individual wagons. As computational power increased, researchers developed one-dimensional multi-mass models. In these models, each wagon and locomotive is represented as a discrete mass, connected to its neighbors by non-linear spring-damper elements representing the draft gears and couplers. The development of these models allowed for the simulation of slack action, the phenomenon where the small amounts of free play in the couplers accumulate, allowing wagons to crash into each other during braking or sudden acceleration. Ouyang and colleagues examine calibrating NTH method for ϕ' in clayey soils using centrifuge CPTu [1]. Significant advancements were made in the mathematical representation of draft gear hysteresis, an essential property where the loading and unloading curves of the draft gear differ, dissipating kinetic energy as heat. Various empirical and semi-empirical models were proposed to capture this behavior, often requiring extensive laboratory testing of draft gear components to derive the necessary coefficients. Despite their sophistication, these physics-based models suffer from a fundamental drawback when applied to operational monitoring: parameter uncertainty. The physical properties of railway components degrade over time due to wear and tear. A draft gear that has been in service for several years will behave very differently from a brand-new one. Since it is economically unfeasible to continuously test and update the parameters of every single wagon in a fleet, physics-based models inevitably diverge from reality over time. Researchers have attempted to mitigate this by implementing adaptive parameter estimation techniques, such as Kalman filters, to update the model states in real-time, but these approaches remain highly sensitive to sensor noise and modeling errors. The self-supervised speech formulation of wav2vec 2.0 suggests a representation-learning comparator for unlabeled sensor sequences, with the acoustic objective redesigned for the physical signals being monitored [2].

2.2 Sensor Technologies and Direct Measurement

The most straightforward method to monitor in-train forces is direct measurement. The railway industry has developed sophisticated sensor technologies to achieve this, most notably instrumented drawgears and instrumented wheelsets. Instrumented drawgears involve installing strain gauges directly onto the coupler shank or the yoke of the draft gear. These gauges measure the microscopic deformation of the metal under load, which can be correlated to the applied force. Instrumented wheelsets go a step further by measuring the forces acting at the wheel-rail interface, providing data on both longitudinal and lateral forces, which is invaluable for predicting derailment risks. Malik and colleagues examine numerical modeling and failure analysis of caisson foundations subjected to the monotonic lateral loads [3]. While direct measurement provides the most accurate ground truth data, it is prohibitively expensive to deploy across an entire fleet of freight wagons. Installing, calibrating, and maintaining these sensors in the harsh railway environment, where they are

ISSN: 2519-9439

exposed to extreme vibrations, dirt, moisture, and massive temperature fluctuations, is a massive logistical challenge. Furthermore, providing a reliable power supply and data transmission system for sensors located on unpowered freight wagons adds another layer of complexity and cost. As a result, direct force measurement is typically restricted to a small number of specialized test trains or specific wagons used for periodic track auditing. The vast majority of commercial freight trains operate without any direct measurement of in-train forces, relying entirely on the driver's skill and the locomotive's onboard systems to ensure safe operation. This gap between the need for comprehensive force monitoring and the economic reality of sensor deployment is the primary driver for the development of indirect, data-driven estimation methods. The literature also considers simultaneous current and vibration measurement based on interferometric fiber optic sensor [4].

2.3 Transition to Data-Driven Methodologies

In recent years, the explosion of artificial intelligence and machine learning has opened new avenues for solving complex engineering problems. In the railway sector, data-driven approaches were initially applied to predictive maintenance and anomaly detection, utilizing historical data to forecast wheel wear, bearing failures, and track degradation. As the algorithms became more robust and capable of handling sequential time-series data, researchers began exploring their application to train dynamics. The core premise of the data-driven approach to in-train force monitoring is that the forces experienced by the train are a deterministic outcome of the operational inputs and the environmental conditions. If a machine learning model is provided with sufficient historical data containing locomotive traction efforts, brake pipe pressures, train speed, track gradients, and the corresponding in-train forces (measured during special test runs), it can learn to predict the forces based purely on the easily measurable operational inputs. Initial studies in this domain utilized relatively simple architectures, such as feedforward artificial neural networks and support vector regression. While these models showed promise in steady-state conditions, they struggled to accurately capture the transient dynamics associated with slack action and rapid brake applications. The temporal dependencies inherent in train dynamics, where the force at the current moment is heavily dependent on the sequence of events leading up to it, require models capable of retaining short-term memory. Consequently, subsequent research shifted towards recurrent neural network architectures and advanced ensemble methods. However, the vast majority of this literature assumes a fixed train configuration for model training and testing. The critical issue of how these models perform when the consist is continuously changed, simulating real-world freight operations, remains significantly underexplored. This paper addresses this specific gap, providing a comprehensive evaluation of data-driven strategies for dynamic consist configurations.

3. Problem Formulation and System Architecture

3.1 Understanding the Physical Phenomenon

To properly formulate the data-driven monitoring problem, it is necessary to thoroughly understand the physical phenomena underlying in-train forces. A heavy-haul train can be conceptualized as a one-dimensional array of coupled masses traversing a three-dimensional track profile. The primary sources of excitation in this system are the traction and dynamic braking efforts applied by the locomotives, the pneumatic or electronically controlled pneumatic braking forces applied at each wagon, and the gravitational forces acting on the consist as it travels over hills and through valleys. When a locomotive applies traction, it pulls the first wagon, generating a draft force in the connecting coupler. This force propagates

ISSN: 2519-9439

sequentially down the length of the train. The speed of this propagation is not instantaneous; it is dictated by the stiffness of the draft gears and the mass of the wagons, functioning much like a mechanical wave traveling through a physical medium. Conversely, when the locomotive applies dynamic braking, it pushes back against the train, generating buff forces. The most dangerous situations arise when different parts of the train are subjected to opposing forces. For example, if the front of the train is cresting a hill and begins to descend, gravity will cause the front wagons to accelerate. If the rear of the train is still ascending the hill, gravity is pulling it backward. This scenario creates massive draft forces in the middle of the train as it threatens to pull itself apart. Another critical element is slack action. Every coupler connection has a small amount of mechanical clearance, or slack, to allow the train to negotiate curves and to enable locomotives to start heavy trains by moving one wagon at a time. In a long train, the cumulative slack can amount to several meters. When a stretched train suddenly brakes, the rear wagons, which have not yet felt the braking force due to pneumatic propagation delays, crash into the slowing front wagons. The slack violently bunches up, generating severe shock forces that can damage cargo and derail wagons. A successful data-driven model must implicitly learn the complex interplay of these physical factors, interpreting how different consist configurations amplify or dampen these effects.

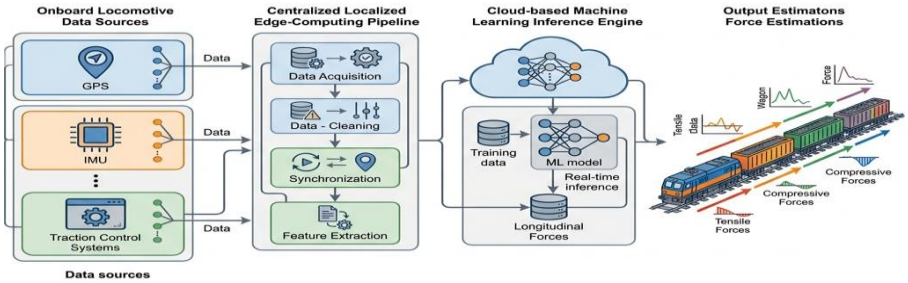


Figure 1Comprehensive Schematic of Decentralized Data

3.2 System Architecture Design

The proposed system architecture for data-driven in-train force monitoring is designed to operate in a dual environment: an offline training and validation phase utilizing historical operational data and high-fidelity simulations, and an online real-time inference phase deployed on locomotive onboard computers. The foundation of the system is the data acquisition module, which continuously records variables from the locomotive control system. Material-agnostic system identification from video [5] suggests a possible complementary observation channel, with visual motion evidence calibrated against the physical state variables inferred by the present system. These variables include the main reservoir pressure, brake pipe pressure, traction motor current, dynamic brake effort, and the physical positions of the driver's control levers. Simultaneously, a global navigation satellite system and an inertial measurement unit provide high-resolution data on the train's speed, acceleration, and geographical location. The geographical location is cross-referenced with an onboard digital

ISSN: 2519-9439

track database. This database provides the look-ahead track topography, giving the system critical information about the upcoming gradients and track curvature that the train is about to traverse. This look-ahead capability is vital, as the forces experienced by the train at any given moment are heavily influenced by the terrain under the entire length of the train, not just the terrain immediately beneath the locomotive. The collected raw data is fed into a preprocessing engine, which is responsible for data alignment, noise filtering, and feature extraction. Crucially, this engine also processes the train consist manifest. The manifest contains the structural DNA of the train for that specific journey, detailing the number of locomotives, the number of wagons, their specific sequence, and the gross weight of each individual vehicle. The central challenge addressed by this architecture is how to mathematically represent this variable-length categorical and numerical data so that a fixed-size machine learning model can process it effectively. The preprocessed feature vectors are then passed to the core machine learning inference engine, which generates continuous estimations of the maximum draft and buff forces occurring within the train, providing early warnings to the driver if predefined safety thresholds are approached.

4. Data Collection and Preprocessing Methodology

4.1 Data Sources and Acquisition

The development of a robust data-driven model relies entirely on the quality, quantity, and diversity of the training data. For this research, a massive dataset was constructed, comprising both empirical data gathered from instrumented revenue-service trains and synthetic data generated using a validated high-fidelity longitudinal train dynamics simulator. The use of synthetic data is a critical strategy in railway research, as it allows for the safe exploration of extreme operational boundary conditions, such as emergency brake applications on steep descending grades, which are too dangerous to stage in the real world. The dataset captures a wide variety of operational scenarios across thousands of kilometers of track. The independent variables, or features, can be broadly categorized into three domains: control inputs, kinematic states, and environmental topography. Control inputs reflect the driver's intentions, including throttle notch positions ranging from idle to maximum power, and brake cylinder pressures. Kinematic states encompass the longitudinal speed and acceleration of the lead locomotive. Environmental topography provides the external boundary conditions, specifically the track gradient and curvature corresponding to the current GPS coordinates. To provide the necessary ground truth targets for model training, the maximum draft force and maximum buff force occurring anywhere in the consist at every time step were recorded. In the physical test trains, these values were derived from wireless instrumented drawgears distributed evenly throughout the consist. In the simulation environment, the forces at every coupler were continuously calculated and the extreme values extracted. Ensuring tight time synchronization between the various sensor streams and the ground truth force measurements is absolutely critical. Even a one-second alignment error between a sudden change in brake pipe pressure and the resulting force peak can severely degrade the learning capacity of the model.

Table 1: Operational and Physical Features Extracted for Model Training

Feature Category	Specific Variables Included	Sampling Frequency	Data Type
Locomotive Control	Traction effort, Dynamic brake effort,	10 Hz	Continuous

ISSN: 2519-9439

	Brake pipe pressure		
Train Kinematics	Longitudinal velocity, Longitudinal acceleration	10 Hz	Continuous
Track Topography	Local gradient, Average gradient under train, Curvature	Spatial (per meter)	Continuous
Consist Configuration	Total mass, Train length, Mass distribution index, Wagon count	Per journey	Static/Derived
Historical Context	Rolling average speed (past 60s), Accumulated brake energy	1 Hz	Derived

4.2 Data Preprocessing and Feature Engineering

Raw sensor data acquired from a locomotive is inherently noisy and subject to intermittent dropouts due to vibrations, electromagnetic interference, and temporary loss of satellite reception. The first stage of preprocessing involves comprehensive data cleaning. Missing values are imputed using linear interpolation for brief dropouts, while longer gaps result in the exclusion of that specific data segment to prevent the introduction of artificial anomalies. High-frequency noise, particularly present in accelerometer and traction current readings, is mitigated using zero-phase low-pass filtering. This ensures that the underlying macroscopic dynamic trends are preserved without introducing phase shifts that would disrupt the temporal alignment of the dataset. Following data cleaning, the continuous variables are normalized. Given that variables like brake pipe pressure operate on a completely different numerical scale than track gradient, normalization is essential to ensure that the machine learning algorithms treat all features with appropriate mathematical weight. Standardization techniques are applied to center the data around a zero mean with a unit standard deviation. Related work has addressed new DMT method for evaluating soil unit weight in soft to firm clays [6]. The most complex aspect of the feature engineering process involves handling the changing consist configurations. A naive approach would be to input the mass of every single wagon in sequence. However, since train lengths vary from a few dozen wagons to over two hundred, this would result in an input vector of highly variable length, which is incompatible with standard machine learning architectures. To solve this, a dimensionality reduction strategy was implemented. Instead of individual wagon masses, the consist is encoded using a set of fixed-length macroscopic descriptors. These include the total mass of the train, the total length, the average mass per wagon, and a calculated mass distribution index. The mass distribution index is a crucial engineered feature; it quantifies whether the mass is concentrated at the front, evenly distributed, or concentrated at the rear of the train. By summarizing the configuration into these representative physical metrics, the models can dynamically adjust their predictions based on the general structural characteristics of the varying consists. The literature also considers evaluating rigidity index, OCR, and su from dilatometer data in soft to firm clays [7].

ISSN: 2519-9439

5. Machine Learning Models for Force Estimation

5.1 Ensemble Tree-Based Approaches

The initial exploration of modeling techniques focused on ensemble tree-based algorithms, specifically Random Forests and gradient boosting frameworks. Random forests supply an ensemble-tree reference for nonlinear prediction and feature-based comparisons [8]. These algorithms are highly regarded in industrial applications due to their robustness against outliers, their ability to handle non-linear feature interactions, and their inherent interpretability through feature importance scoring. The Random Forest algorithm constructs a multitude of independent decision trees during the training phase. Each tree is trained on a random subset of the training data and a random subset of the available features. When making a prediction, the output of all individual trees is averaged. This bagging technique significantly reduces the variance of the model, preventing it from memorizing the noise in the training data, a common problem known as overfitting. In the context of train dynamics, the decision trees learn to partition the operational space based on critical thresholds. For instance, a tree might learn that if the track gradient is steeply negative and the train speed is high, a sudden application of pneumatic brakes will likely result in high buff forces, mapping this logical sequence to a specific force value output. Gradient boosting algorithms take a slightly different approach. Gradient boosting constructs an additive predictor through successive loss reduction, providing a standard supervised comparator [9]. Instead of training independent trees, they train trees sequentially. Each new tree is specifically designed to correct the residual errors made by the combination of all previously trained trees. This iterative boosting process creates a highly accurate predictive model. While these tree-based models excel at capturing complex, static non-linear relationships, their standard implementations treat every time step as an independent event. They do not inherently possess a mechanism to remember past states. To apply them to the time-series problem of train dynamics, the temporal context must be manually engineered into the feature space. This is achieved through sliding window techniques, where the input vector at the current time step includes the sensor readings from several previous time steps. While functional, this manual windowing significantly increases the dimensionality of the input space and can struggle to capture very long-term dependencies, such as the slow recharge of the pneumatic brake pipe over several minutes.

5.2 Recurrent Neural Networks for Sequential Dynamics

To overcome the limitations of static models, the research shifted towards deep learning architectures specifically designed for sequential data processing, namely Recurrent Neural Networks. Unlike traditional feedforward networks where data flows strictly in one direction from input to output, recurrent networks contain internal loops. These loops allow information to persist, meaning the network maintains an internal hidden state that acts as a memory of past inputs. When processing the data for the current time step, the network utilizes both the current operational features and its internal memory of the previous time steps, making it theoretically ideal for modeling the continuous physical evolution of train dynamics. However, standard recurrent networks suffer from the vanishing gradient problem, where they lose the ability to learn connections between events separated by long time delays. To address this, specialized architectures known as Long Short-Term Memory networks were employed. These networks introduce a complex internal cellular structure governed by distinct regulatory mechanisms referred to as gates. These gates control the flow of information, explicitly deciding what old information is irrelevant and should be forgotten, what new information is important and should be stored, and what portion of the internal memory is relevant for generating the output at the current moment.

ISSN: 2519-9439

In the application of predicting in-train forces, this architecture allows the model to continuously track the physical state of the train. For example, if the locomotive applies traction, the model registers the input and updates its internal state to reflect that the train is now in a stretched condition. If the train subsequently encounters a downhill gradient ten minutes later and applies brakes, the model utilizes its retained knowledge of the stretched state to accurately predict the ensuing slack run-in forces. The ability to autonomously maintain this internal representation of the train's dynamic state, without relying on manual feature windowing, represents a significant leap forward in modeling complex mechanical systems.

6. Experimental Results and Discussion

6.1 Evaluation Methodology and Metrics

The performance of the developed machine learning models was evaluated using a rigorous hold-out validation strategy. The dataset was partitioned, ensuring that the models were tested on entirely distinct journeys and consist configurations that they had never encountered during the training phase. This guarantees that the evaluation reflects the models' true generalization capabilities in a real-world setting where consist variations are the norm. Another relevant line of investigation concerns intelligent safety evaluation of tunnel lining cracks based on machine learning [10]. The predictive accuracy was quantified using standard statistical regression metrics. The primary metric utilized was the Mean Absolute Error, which provides a straightforward interpretation of the average magnitude of the prediction errors without considering their direction. However, in the context of railway safety, predicting large peak forces accurately is far more critical than predicting low-level nominal forces. Missing a massive force spike could result in a derailment, whereas small errors during steady cruising are operationally irrelevant. To account for this, the Root Mean Square Error was heavily scrutinized. By squaring the errors before averaging them, this metric heavily penalizes large deviations from the ground truth, making it an excellent indicator of a model's reliability in extreme boundary conditions.

Table 2Comparative Performance of Predictive Models on Unseen Consist Configurations

Model Architecture	Mean Absolute Error (kN)	Root Mean Square Error (kN)	99th Percentile Error (kN)
Baseline Physical Model	145.2	210.5	340.1
Random Forest (Sliding Window)	85.6	134.2	215.8
Gradient Boosting (XGBoost)	78.3	126.7	198.4
Long Short-Term Memory (LSTM)	42.1	68.5	110.3

6.2 Analysis of Predictive Performance

The experimental results demonstrate a clear hierarchy in the performance of the various modeling approaches. The baseline physical model, populated with generic default parameters due to the lack of specific measurements for the varied consists, exhibited the

ISSN: 2519-9439

highest error rates. Its inability to dynamically adjust to the changing friction coefficients and draft gear wear rates resulted in significant deviations, particularly during transient braking events where slack action occurs. The ensemble tree-based models, utilizing manually engineered sliding windows for temporal context, showed a massive improvement over the static physical model. By identifying complex patterns in the historical operational data, these models successfully anticipated general force trends. However, their performance degraded during highly dynamic events characterized by rapid changes in acceleration. The manual windowing technique, while providing some historical context, proved insufficient for capturing the deep, complex temporal dependencies of air brake propagation and sustained mechanical oscillations. The Long Short-Term Memory network outperformed all other approaches by a substantial margin. The error metrics indicate a profound improvement in accuracy, effectively halving the errors observed in the tree-based models. The true strength of the sequential architecture became apparent during the analysis of peak force events. When a heavy-haul train transitions over a cresting grade while simultaneously manipulating the throttle, the resulting wave dynamics are extremely complex. The deep learning model successfully maintained an accurate internal representation of the train's changing physical state throughout these maneuvers. It demonstrated an exceptional ability to predict the timing and magnitude of severe run-in and run-out shocks, proving that its internal gating mechanisms successfully learned the underlying physical principles governing wave propagation and slack accumulation purely from the data streams.

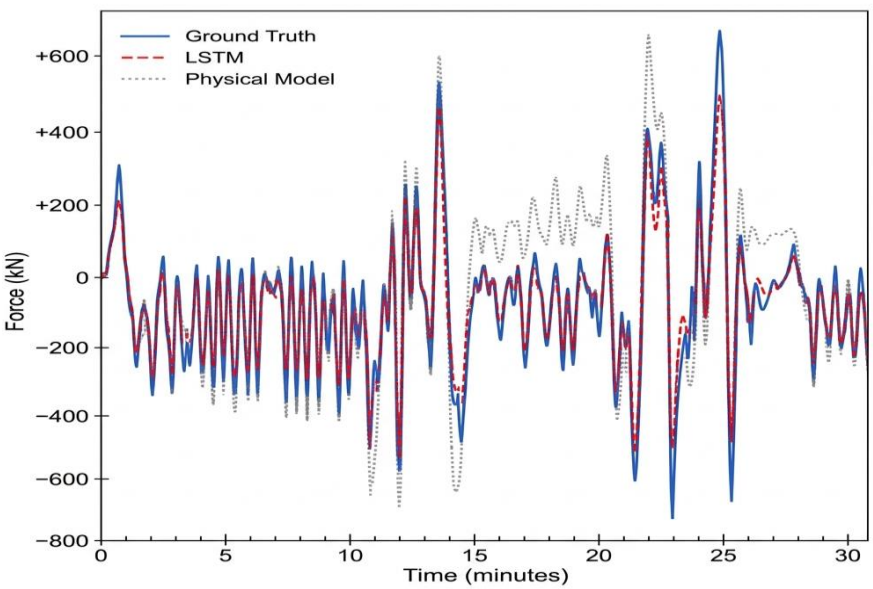


Figure 2Comparative Timeseries Graph of In

6.3 Operational Implications and Edge Cases

Beyond the pure statistical metrics, it is vital to analyze the models' behavior in specific operational edge cases. One critical scenario evaluated was a sudden, severe application of the dynamic brake while operating on a fluctuating, undulating track profile. Under these conditions, the distribution of mass within the consist plays a dominant role. When evaluated on a consist with a heavy concentration of loaded wagons at the rear pushing against empty wagons at the front, the sequential model accurately predicted a dangerous surge in compressive buff forces, correctly identifying the high risk of derailment. These results confirm that the macroscopic feature engineering strategy, summarizing the consist into mass distribution indices and length parameters, was highly successful. The model did not need to

ISSN: 2519-9439

know the specific mass of wagon number seventy-two; it only needed to know that the rear third of the train was significantly heavier than the front. This generalization capability is what makes the proposed system viable for real-world deployment. Instead of requiring exhaustive configuration protocols and calibration runs every time a train is assembled, the system can instantaneously adapt to the new manifest data, providing reliable, real-time safety monitoring from the moment the locomotive departs the shunting yard.

7. Conclusion and Future Work

7.1 Summary of Findings

The continuous and reliable monitoring of in-train forces remains an indispensable requirement for the safe expansion and optimization of heavy-haul freight railway networks. This comprehensive study investigated the transition from classical, static physics-based simulations to dynamic, data-driven methodologies, specifically addressing the persistent operational challenge of continuously changing consist configurations and payload distributions. By formulating the problem around the complex mechanical realities of slack action, draft gear hysteresis, and distributed train mass, a robust, decentralized data architecture was proposed and evaluated. The empirical analysis conclusively demonstrated that machine learning algorithms, particularly deep learning architectures designed for sequential time-series analysis, are highly capable of learning the complex, non-linear dynamics of multi-kilometer freight trains. The implementation of advanced recurrent structures successfully captured the deep temporal dependencies inherent in pneumatic braking and mechanical wave propagation. Crucially, the utilization of macroscopic consist descriptors proved to be an effective strategy for handling variable-length train manifests, allowing the predictive models to generalize accurately across entirely novel train configurations. The deep learning approach significantly outperformed traditional methods, particularly in its ability to predict critical peak forces during extreme transient maneuvers, highlighting its potential to dramatically enhance operational safety.

7.2 Future Research Directions

While the results presented in this study establish a strong foundation for data-driven force monitoring, several avenues for future research remain. The current architecture relies on centralized processing or powerful onboard computing systems. Future work should explore the optimization and quantization of these deep neural networks for deployment on resource-constrained edge computing devices directly within the locomotive control systems. Furthermore, integrating these predictive models into a closed-loop automated train operation system represents the ultimate goal. By allowing the artificial intelligence not just to monitor forces, but to actively manipulate the throttle and brakes in real-time to proactively minimize those forces, railway operators could achieve unprecedented levels of safety, energy efficiency, and network throughput. Finally, expanding the data acquisition to include real-time localized weather data and track adhesion estimations could further refine the models' accuracy under adverse environmental conditions, pushing the boundaries of intelligent railway operations.

References

[1] Ouyang, Z., & Mayne, P. W. (2018). Calibrating NTH method for φ′ in clayey soils using centrifuge CPTu. In Cone Penetration Testing 2018 (pp. 491-497). CRC Press.

ISSN: 2519-9439

[2] Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. In Advances in Neural Information Processing Systems.

[3] Malik, B. A., Mandhaniya, P., Soni, A. K., & Ouyang, Z. (2024). Numerical modeling and failure analysis of caisson foundations subjected to the monotonic lateral loads. Ocean Engineering, 309, 118491.

[4] Yu A, Pang F, Yuan Y, et al. Simultaneous current and vibration measurement based on interferometric fiber optic sensor[J]. Optics & Laser Technology, 2023, 161: 109223.

[5] Zhao, Y., Chen, H., Liu, C., Li, Z., Herrmann, C., Hur, J., ... & Xu, M. (2025, October). Toward material-agnostic system identification from videos. In 2025 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 5944-5956). IEEE.

[6] Ouyang, Z., & Mayne, P. W. (2016). New DMT method for evaluating soil unit weight in soft to firm clays. Geotechnical and Geophysical Site Characterisation; Lehane, B., Acosta-Martínez, HE, Kelly, R., Eds, 785-790.

[7] López-Acosta, N. P. (2019, November). Evaluating rigidity index, OCR, and su from dilatometer data in soft to firm clays. In Geotechnical Engineering in the XXI Century: Lessons learned and future challenges: Proceedings of the XVI Pan-American Conference on Soil Mechanics and Geotechnical Engineering (XVI PCSMGE), 17-20 November 2019, Cancun, Mexico (p. 286). IOS Press.

[8] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[9] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

[10] Jin, X., Qiao, K., Bu, M., Wang, J., Wang, M., & Fang, C. (2025). Intelligent safety evaluation of tunnel lining cracks based on machine learning. Engineering Failure Analysis, 167, 109082.