Intelligent Health State Awareness System Framework for Maintenance-Free Railway System Vision

Dy D. Le, Director
Texas Tech University, Office of the Vice President for Research
Institute for Materials, Manufacturing, and Sustainment
Lubbock, TX 79409

Abstract

To increase safety significantly and make the train and its track infrastructure substantially less expensive to operate and maintain, a paradigm shift with respect to the overall operational sustainment is needed. Additionally, to reduce train accidents and enable the railway system maintenance-free vision, an artificial intelligence (AI)-machine learning-based monitoring technology, which is capable of autonomously providing the health state awareness of the entire wheel-track system and direct feedbacks for adaptive positive train control (PTC), is needed. Automated health state awareness and adaptive PTC system cannot be effectively designed and implemented without the integration of AI-related technologies including machine learning, which are computer intensive and highly parallelized. As new central processing unit architectures customized for AI applications have been evolving at exponential speeds, the accurate and critical complex mechanical system's health state awareness information, extracted from heterogeneous sources, can be autonomously derived with the use of the machine learning in real-time or near real-time environment. As a result, the health state awareness technology can now be developed to enable the automation of data fusion and analysis, minimization of required data for trending and analysis, and more effective operation and maintenance of trains and track infrastructures. This paper outlines a framework for a machine learning-based health state awareness system to address the railway safety and provide potentials for achieving the maintenance-free vision.

Key-words

Artificial intelligent, machine learning, adaptive positive train control, prognostics, diagnostics, maintenance-free, railway system

1.0 INTRODUCTION

The year of 2018 continues to present passengers and operators with depressing fatal train derailment and collision news. Whether the train derailments and collisions were caused by mechanical wheel-rail unmatched dynamic interactions; mechanical component failures; poor maintenance; or human errors, these accidents have caused severe safety and financial concerns and detrimental impacts on the public transportation perception.

Train accidents can mostly be caused by derailments in which track and equipment failure are the primary cause of accident on the main tracks [1]. Additionally, according to the U.S. Department of Transportation (USDOT) Volpe Center-Innovation Research Program Office, 70% of railroad derailments were reported in the U.S. in 2016 [2]. As a result, the Federal Railroad Administration (FRA) is pursuing numerous research efforts to monitor the critical behavior of the wheel-track interactions to assess adverse conditions of the

Second International Workshop on Structural Health Monitoring for Railway System, Qingdao, China, 17-19 October 2018

wheel and/or track that may affect the train safety and performance.

Monitoring the behavior of the train and railway infrastructure elements is a complex task, particularly, when it is desirable to conduct it in real time. There is a wide range of sophisticated diagnostics systems that have been used to support the health monitoring of the wheel/track systems. These systems include track, overhead line, and vehicle dynamics measurement as well as vision system, video inspection, and other capabilities such as signal processing, railroad infrastructure kinematic analysis, and environmental temperature measurement. In fact, using those monitoring and measurement technologies, the railroad industry has been collecting a tremendous amount of data, and the concept of "big data" has become one of highly complicated research topics within the railroad community for the last decade. Consequently, the research involved in "big data" with the railroad community has taken a big step in using the artificial intelligence (AI) and machine learning technology for the data mining and correlation tasks to potentially differentiate anomalies that may exist in complex patterns. With the advancement of the AI machine learning and the amount of data, which has already been collected, the railroad industry learns many techniques for fine tuning the data collection, processing, and reporting as well as visualization. The developed AI-machine learning technologies and their applications can potentially relieve the railroad maintainers and analysts from the inherent burden with time intensive tasks associated with engineering and analysis of a vast amount of data collected using monitoring methodologies. This "big data" analysis burden, if not addressed, can limit the U.S. railroad operational effectiveness, impede their safety improvement, and make them expensive to operate and maintain.

To increase safety significantly and make the train and its track infrastructure substantially less expensive to operate and maintain, a paradigm shift with respect to the overall operational sustainment is needed. Additionally, to reduce train accidents and enable the railway system maintenance-free vision, an artificial intelligence (AI)-machine learning-based monitoring technology, which is capable of autonomously providing the health state awareness of the entire wheel-track system and direct feedbacks for adaptive positive train control (PTC), is needed. Automated health state awareness and adaptive PTC system cannot be effectively designed and implemented without the integration of AI-related technologies including machine learning, which are computer intensive and highly parallelized. As new central processing unit architectures customized for AI applications have been evolving at exponential speeds, the accurate and critical complex mechanical system's health state awareness information, extracted from heterogeneous sources, can be autonomously derived with the use of the machine learning in real-time or near real-time environment. As a result, the health state awareness technology can now be developed to enable the automation of data fusion and analysis, minimization of required data for trending and analysis, and more effective operation and maintenance of trains and track infrastructures. With appropriate physics-based models and intelligent adaptive controls, the health state awareness technology can also provide an automation capability to achieve the maintenance-free vision for the railway systems. The maintenance-free vision for the mechanical systems has been pursued for quite a long time. The U.S. Army, for example, is currently planning to develop the next-generation helicopters with "fatigue-free" and "zero-maintenance" characteristics [3].

2.0 INTELLIGENT DIAGNOSTICS AND CONTROLS FRAMEWORK

A framework has been developed for a proof-of-concept design for a machine learning-based health state awareness system that can be potentially installed on a rolling stock for testing and demonstrating its innovation and effectiveness. The methodology hereby called "intelligent Diagnostics and Controls", or iDaC, is aimed at enabling the real-time monitoring and situation response applicable to railroad and train platforms.

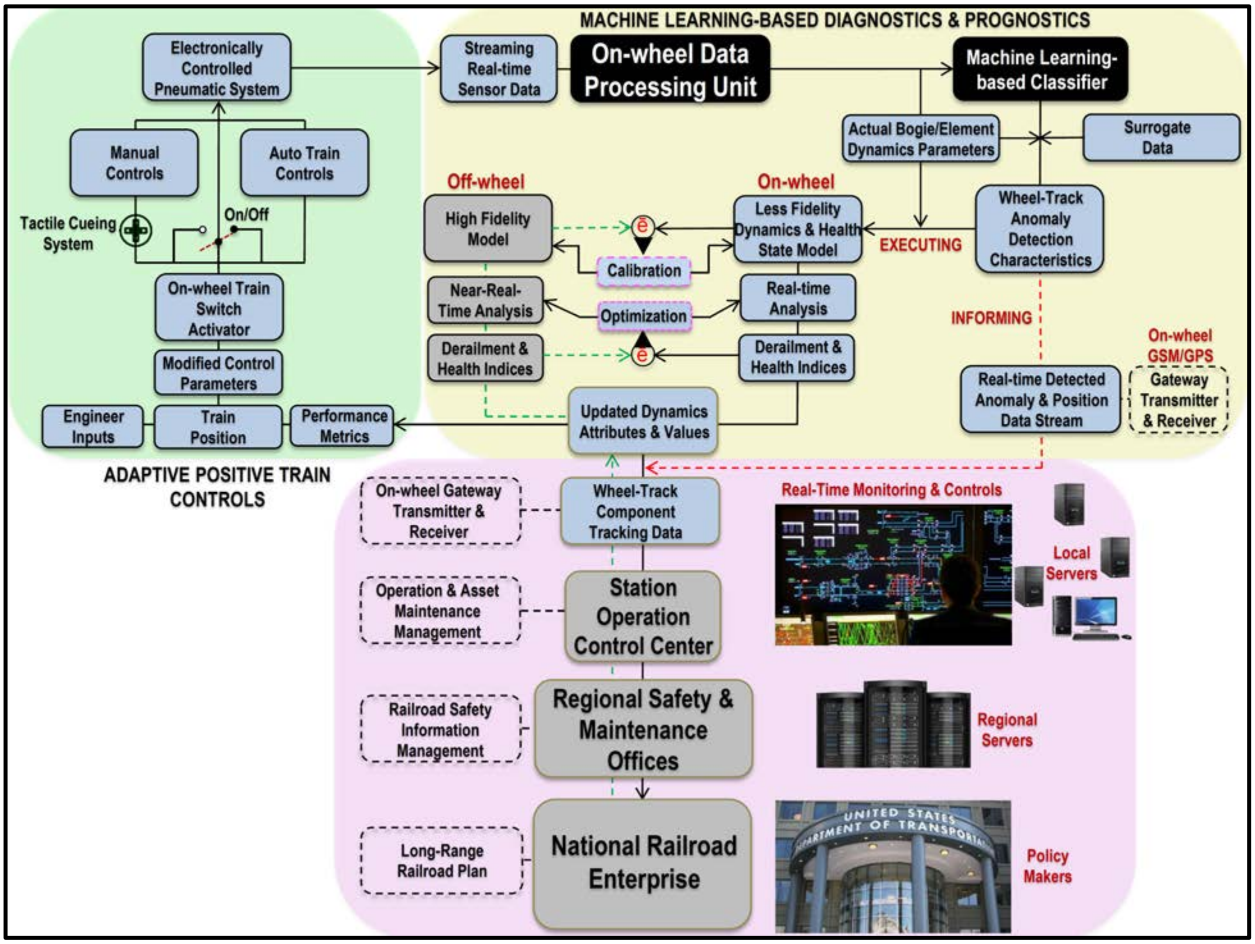

Figure 1: Overall Machine Learning-based Diagnostics and Prognostics Concept

Particularly, the conceptualized system framework is focused on the vehicle-track interactions, but can be modularly reconfigured to extract relevant data from a variety of existing monitoring systems including track, overhead line, and vehicle dynamics measurement as well as vision system, video inspection, and other sources such as signal processing, railroad infrastructure kinematic analysis, and environmental temperature measurements.

The iDac system framework is envisioned to be reconfigurable to have the following automated operational functionalities:

- Ingest and transform real-time unsynchronized, unstructured/structured data from heterogeneous sources (e.g., sensor signals, on-board monitoring and control system, off-board inspections, relevant information from maintenance/inspection databases) into recognized patterns.
- Using defined machine learning-based techniques, analyze patterns to detect anomalies (e.g., abnormal rolling vibration, vehicle dynamic unstable condition, derailment and failure precursor, fatigue crack, and operational damage) and assess the train decreased performance and

mechanical element integrity (e.g., extreme angle of attack, inter-axle misalignment, hunting severity, poor brake performance, and defective components).

- Identify a set of influential operating factors (e.g., derailment coefficient, lateral/vertical contact force, wheel load unbalance ratio, damping, critical speed, and track curving radius) and develop modified control parameters (e.g., speed and braking), which are autonomously fed into the PTC feedback loop to prevent derailment and other related catastrophic failures. When the identified influential operating factors combined with or integrated into appropriate physics-models, automation strategies can be potentially developed to reduce the maintenance frequency to minimum or zero within a desirable maintenance-free interval.
- Provide platforms and operators options to manually or autonomously execute speed control authority accordingly to: (a) prevent derailment; decelerate or stop crack/damage growth; and mitigate other detected anomalies, and (b) optimize in-situ operational performance.

To achieve the above capabilities, combined approaches, which integrate the state-of-the-art, commercial-of-the-shelf (COTS) high performance neural enabled hardware and novel machine learning algorithms to provide the vehicle-track health state awareness in real-time or near-real-time, are needed. It is envisioned that the iDaC system framework, with future interface upgrades and modifications, will be capable of extracting sensor signals from many other diverse heterogeneous sources for automated data analytics and analysis using the machine learning techniques. Figure 1 illustrates the overall concept framework, where its functionalities can be extended to be integrated into existing or next-generation train platforms with PTC to increase safety and reduce operation and sustainment costs substantially.

2.1 CURRENT STATE-OF-THE-ART MACHINE LEARNING

Machine learning with the use of the deep neural network is becoming a powerful tool for any aspect of modern systems involving health care, business, social issues, engineering etc. Regardless of what business sector we are in, the amount of data collected has become astronomically large to the point that human is no longer able to manage the manually-extensive and time-consuming analysis.

For more than a decade, FRA has been involved in the research and development efforts including AI neural networks and machine learning to hopefully bring automation to inspection and maintenance as well as train operations and performance monitoring. However, many research and development efforts, particularly in the machine learning, so far have been mainly theoretical and not yet been fully matured for practical applications in the railroad operation and sustainment. As the railroad industry has also been accelerating in collecting a vast amount of data on railroad behaviors including the wheel-track interactions, a strong demand exists for a machine learning capability to automate the data mining process and information discovery to provide relevant information for making intelligent maintenance and safety-related decisions.

In the railroad maintenance area, numerous research in the past showed that the AI-machine learning could predict the rate of rail wear by learning the characteristics of the rail track including degree of curvature, deviation of curvature, superelevation, and lubrication etc. For the railroad operational situation awareness, the AI technology can analyze real-time data from the running trains to provide real-time locations of trains and their respective systems and sub-systems in each train, insight into the operational performance of trains, and inter-relationship of various systems across the entire railroad fleet.

However, today there is no matured machine learning-based technology that can be integrated into trains to provide the health state awareness with direct feedbacks to the PTC system for appropriate speed controls to prevent catastrophic failures or collisions.

2.2 DATA STREAMING

As we are facing with interoperability issues, the iDaC framework takes advantage of matured open-source technologies and widely used standards as building blocks for the future innovation and technology advancement and easy upgrades at low costs.

Additionally, complex systems including trains, with existing embedded sensors; on-board monitoring devices; train integrated management system; and data buses, provide operators with large volumes of heterogeneous data. To achieve the commonality goal and address data heterogeneity issues, the iDaC framework integrates key open-source methods to enable the automated data streaming, ingestion, aggregation, and transformation while applying the machine learning and knowledge discovery from data (KDD) techniques to enable in-memory computation and provide the necessary diagnostics and prognostics.

To enable the real-time or near real-time state awareness and adaptive positive train controls, a continuous streaming of data generated by embedded sensors or other monitoring systems on-board of trains is required. This can be done by tapping into the appropriate data buses and interfaces used in the existing sensing and monitoring systems, many of which are capable of converting the analog signal inputs, generated from embedded sensors and transducers, into digital forms. Once sensor signal inputs are in digital forms, selected data stream processing techniques are used to move large volumes of data from different sources on-board of trains, which provide the iDaC with continuously updated state of complex systems and data while trains are in operation.

The envisioned iDaC suite of data stream processing techniques is built using open-source applications including a data handling tool, to convert unstructured data into structured and standardized formats containing data models and content characteristics (metadata). Specified in the data handling tool configuration file is a list of possible data formats and schema, e.g., structure of an observation and data type of each field of an event. Data formats, used to describe structured data and serialize objects, include Column-Separate Value (CSV), JavaScript Object Notation (JSON), and Extensible Markup Language (XML). XML format is Standard Generalized Markup Language (SGML) compliant, e.g., independent of any proprietary software, and widely used in many applications. The data handling tool extracts and performs formatting on the incoming data stream in real-time or near real-time and provides data files readable by other applications integrated within the iDaC data streaming and pipeline architecture.

Apache Kafka is a distributed streaming platform used to build real-time streaming data pipeline and allow reliable access of processed data between applications in the envisioned "IDaC" framework. Kafka is also used to build real-time streaming applications that can transform or react to the stream of data. Kafka allows storing and processing of historical data from the past and real-time data as it arrives. It is a scalable publish-subscribe messaging system that stores streams of *Records* in categories called *topics* and is consisted of four components or Application Program Interfaces (APIs): Producer, Consumer, Stream, and Connector APIs. Through Kafka, multiple applications (*Consumer*) can consume *Topics* (processed data) independently to perform data mining and analytics or machine learning techniques [4].

To enable the real-time or near real-time distributed processing of incoming data, Apache Storm is used to analyze smaller clusters of data, which are dynamically changing or being digested into the data pipeline at high velocity. Apache Storm enables the capability for in-memory computing, which allows the application to move working datasets and perform calculations within the computer Random Access Memory or RAM. Apache Storm forms a topology, which takes input streams or *Spouts* and processes them to provide output modules called *Bolts* [5].

As the train platform sensor data streams come in and are processed by the data handling tool, the processed data are collected and handled by Apache Kafka cluster where *Spouts* are tasked to read or subscribe to *Topics* and then push the sensor data into the *Data Preprocessing Bolt*. Within the Storm cluster, *Bolt* preprocessing operations including data transformation and filtering are performed on the sensor data and then send it to the *Data Mining Bolt* for performing prediction of outcome using the fault classifier developed and trained from historical sensor data.

2.3 MACHINE LEARNING-BASED DIAGNOSTIC AND PROGNOSTIC METHODS

Train-Rail Interaction Models and Monitoring Parameters: For the machine learning training scheme and comprehensive vehicle dynamics analysis for the on-board hybrid intelligent system in real-time, a coupled multi-body wheel-rail interaction physics-based model is developed to obtain the dynamic response and performance of the train interacting with the inner/outer rails. The modeling execution is automated (e.g., to receive data inputs from established on-board sensing systems) to predict or project critical wheel-rail dynamic parameters in real time. These computed dynamic parameters are autonomously compared with established threshold values, which, if exceeded, will have a high probability of causing catastrophic accidents or component failures. Additionally, the computed parameters and dynamic information can also be autonomously fused to provide appropriate health indices, which are visible on the train dashboard for the train comprehensive health state awareness in real time.

There are numerous wheel-rail dynamic models, which have been successfully developed by the railroad industry and academic community. A thorough study needs to be conducted to select appropriate sub-models, which have already been verified and validated. The selected wheel-rail interaction sub-models can then be modified and enhanced for coupling and real-time computation and analysis. Additionally, to accommodate the real-time computation, (e.g., less number of degrees of freedom, fast convergence, and high degree of accuracy), selected techniques such as cross iteration algorithm and relaxation method are needed for developing model algorithms [6].

To monitor the interaction of the train components and track and predict the vibration characteristics, a coupled multi-body wheel-rail interaction physics-based model, as mentioned above, is required. This model consists of several sub-models coupled together to characterize the lateral, vertical, and longitudinal dynamics behavior of the entire train (e.g., several train wagons and bogies). Additionally, the track and its substructures (e.g., rails, sleepers, and ballasts) are modeled as elastic beams and coupled with the train dynamic sub-models to provide a whole wheel-rail system response and behavior. The coupled wheel-rail interaction physics-based model is integrated with reduced-order mathematical equations for the real-time execution of modeling to track the dynamics of wheel-rail interactions and behavior white the train is passing through tracks.

Sub-models, which are coupled together, address not only the performance and vibration characteristics of the wheel and rail in lateral, vertical, and longitudinal directions, but also the effect of braking system. The lateral dynamic model provides the vehicle dynamic behavior in the lateral displacement, roll, and yaw directions and is used to monitor the lateral stability including hunting, rail curving effect, and derailment. The vertical dynamic model provides the dynamic performance of suspension systems and wheel-rail as a whole. The longitudinal dynamic model can be used to monitor the performance of the train braking system on the dynamics of the whole train and potential derailment as the train is running through curve tracks.

There are numerous critical dynamic parameters, produced by these models, which can be monitored while trains are running to ensure passenger safety and performance of the wheel-rail system. Critical dynamic parameters and data, that may affect the safety and performance, are:

a. Maximum train vibrations induced by critical speed (e.g., natural frequency), which generates the resonance and instability of the wheel-rail system. When the train speed (TS) reaches a critical speed (CS), the amplitude of the train vibration grows exponentially with time.

b. Derailment precursors can be predicted using the wheel-rail contact forces while trains are running on track. From the wheel-rail contact forces, the derailment coefficient (DC), which depends on many factors including track curve and lubrication/friction coefficient, can be computed or estimated. When the train running DC exceeds a safety limit, derailment can happen. Additionally, using the DC data and information generated by trains running on track in real time, maintenance actions can be planned for wayside lubrication on tracks to reduce the potential derailment effect.

c. For those train bogies designed with semi-active suspension controls, the damping effect can change in real time. When the damping ratio (DR) reduction exceeds a threshold value, the train can reach its critical speed and become unstable. So, the influential relationship of critical speeds of trains and suspension stiffness and damping (SD) can be monitored and/or adjusted to ensure safety and better performance.

d. The maximum and total amplitudes of acceleration for the rail and sub-structures, bogies, primary/secondary suspension, and wheels increase with the increase of the train speed, which also has great influence on the wheel-to-rail contact forces.

e. Wheel bearing temperature measurements.

Train-Rail Machine Learning-based System Functionalities. Figure 1 illustrates a framework for the machine learning-based diagnostics, prognostics, and adaptive controls designed to prevent the train catastrophic events. As the raw sensor data and information are being streamed, the envisioned machine learning-based system (e.g., hardware/software), when installed on trains, can perform these essential tasks in real time:

a) Execute the coupled multi-body wheel-rail interaction physics-based model (low fidelity): During the execution of the modeling and analysis, critical parameters based on actual operational data, are computed, derived, and consequently, fused into health condition indicators or indices. The computation and analysis are designed to be in real time with low fidelity models, but calibrated and optimized with off-wheel modeling solutions in near real-time using on-board gateway transmitter and receiver.

b) Conduct machine-learning-based fault classifications: While the wheel-rail interaction computation is being done, the system simultaneously conducts the fault classification using the machine learning-based training models, which will be described in detail in the next section. If an anomaly were detected, the relevant wheel-track anomaly characteristics would be transmitted immediately to the train operation stakeholders including the notional Station Operation Control Center, Regional Safety and Maintenance Offices, and potentially National Railroad Enterprise. At a particular train station, a local high performing server system is also designed to receive real-time inputs from the running trains to perform the coupled multi-body wheel-rail interaction physics-based model but with significantly higher fidelity and more complicated neural network (more hidden layers). The results would be then automatically transmitted to the running train on-board server system for calibration and optimization with real-time solutions.

c) Modify influential parameters: Upon having computed and derived the health condition indices and/or fault characteristics, the on-wheel system will modify influential parameters (e.g., speed, braking, and/or other operational parameters), which have direct dynamic relationship between, for example, derailment and fault acceleration and provide direct inputs to the on-board adaptive PTC system. The notional adaptive PTC system, installed on trains, can have an on/off switch for train controls. If an "off" position is selected, upon having received modified parameters/inputs, the system can providing sensitive cueing signals and/or visual/audio warnings for operator's prompt actions. If an "on" position is selected, the adaptive PTC system can take over the controls and immediately execute the appropriate actions to reduce the speed or stop the train completely. The notional Station Operation Control Center can also perform the envisioning adaptive PTC functionality, if adaptive PTC is not available on trains.

Machine Learning-based Diagnostics and Prognostics Method. There has been a number of machine learning methods developed for various applications including fault diagnostics

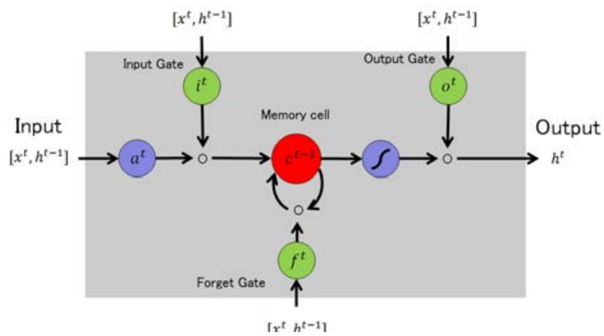

Figure 2: LSTM Cell

and prognostics for many mechanical platforms. Taking advantage of the work, which has been substantially investigated, the iDaC framework can extend the Long Short-Term Memory (LSTM) method, as described in [7] and integrates it with Auto-Encoder and Decoder (AE/AD) models (e.g., to classify the train dynamic instability, derailment precursors, corrugated rail severity or irregularities, and other potential faults).

LSTM-AE/AD has been used in many applications including multi-sensor prognostics for the aviation platform systems [8-9]. LSTM is a machine learning method for sequence tasks, and its cells contain information outside the normal flow of the recurrent network in a gated cell, in which information can be stored in, written to, or read from. LSTM is a suited neural network for sequential sensor data and for discovering hidden patterns from multiple operation data and fault models developed from the sensor networks installed on-board of mechanical platforms. An LSTM cell makes decisions about what to store, and when to allow reads, writes and erasures, via gates that open and close. A typical LSTM cell is shown in Figure 2. It uses input x^t and hidden state activation h^{t-1}, and memory cell activation c^{t-1} to compute the hidden state activation c^t at time t. LSTM cell uses the Input Gate i^t, Forget Gate f^t, and Output Gate o^t, to make a decision if: (1) an input needs to be remembered, (2) a previous memory needs to be retained, and (3) when a memory content needs to be output (h^t). Detail explanations of the LSTM neural network can be found in [7].

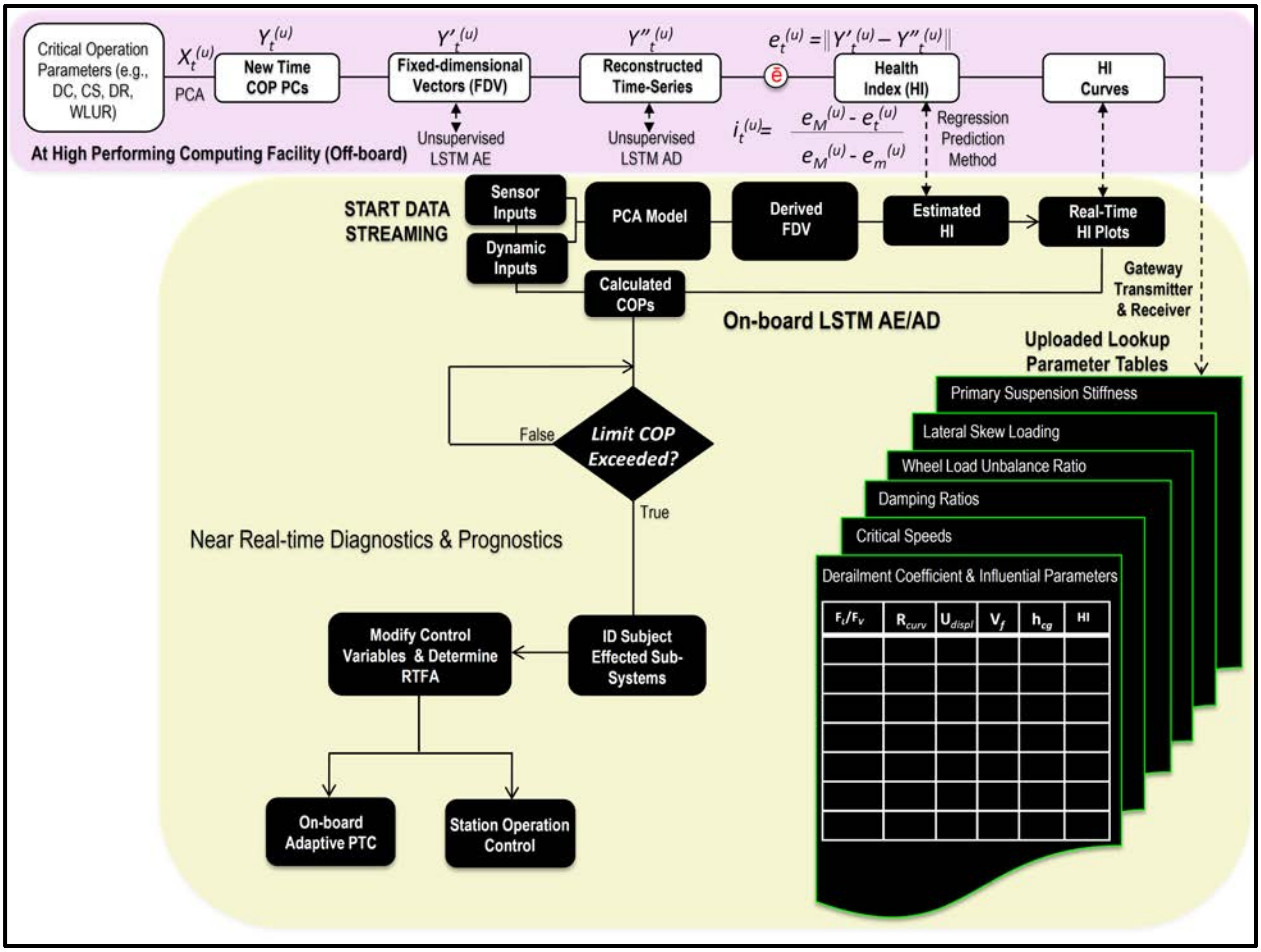

Figure 3: Train-Rail Machine Learning-based Method

Figure 3 depicts a framework of classifying potential faults, which have direct relationships with critical operation parameters or COPs (e.g., derailment coefficient (DC), critical speed (CS), damping ratio (DR), wheel load unbalance ratio (WLUR)). The mentioned parameters are a few examples, and more influential COPs can be defined. In the envisioned machine learning approach, LSTM-AE/AD models can be continuously trained using generated data from historical accident analyses and/or results from the modeling and simulation including using the original coupled multi-body wheel-rail interaction physics-based model. Using the Principal Component (PC) Analysis (PCA) technique, we transform the historical or simulated data/information with and without faults or anomaly into new time models and principal components (PCs) with minimized linear relationship among data $Y_t^{(u)}$. LSTM-AE then maps these input sequences to a fixed-dimensional vector (FDV) representation $Y'^{(u)}_t$. LSTM-AE then tries to learn abstract of the $Y'^{(u)}_t$

function and estimate the same input patterns as its output. LSTM-AD learns from this FDV model to reconstruct the time-series of the train model in reverse order using the predicted values from the last time step in the current hidden LSTM layer $Y''^{(u)}_t$. The error ($e_t^{(u)}$) presented at each step is minimized, until a desirable error set point is met (e.g., 1%), normalized, and used to establish the Health Index (HI) for each event, $i_t^{(u)}$, in which $e_M^{(u)}$ is the maximum and $e_m^{(u)}$ minimum values of reconstruction errors for an event u over a time period t. Using HIs, curves of HIs for each sub-system can be plotted and its required time for action (RTFA) can be derived.

To obtain real-time sub-system-specific $i_t^{(u)}$ and HI curves for estimating RFTA, trained weighting and errors and standard linear regression-based index estimation are applied to the on-board LSTM AE/AD neural net (with fewer hidden layers) using extracted data from actual operational conditions and matching them with the trained HI curves. Healthy and faulty

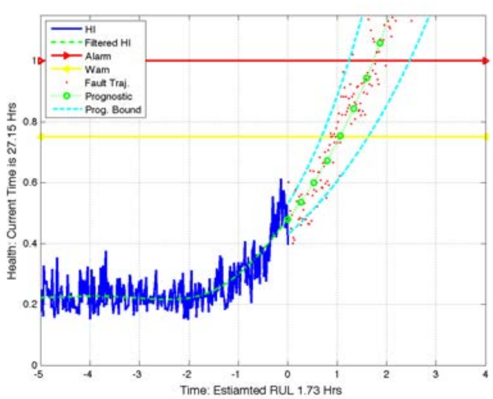

Figure 4: Health Indicator Curve

sub-system may have a value from 0 and 1. Continued operations with HI of 1 or greater can cause catastrophic failures. HIs have been used successfully on many mechanical system monitoring applications including aviation. A typical Health Indicator versus flight cycles curve is shown in Figure 4 [10]. A modified HI curve will be developed to provide the health indication for train-track systems. RFTA can be derived by using the HI predicted time to catastrophic event, if fault were detected and trains would continue to run at the current operational instance.

2.4 MODIFIED CONTROL PARAMETERS FOR NOTIONAL ADAPTIVE PTC

The envisioned iDaC framework, with appropriate interfaces communicating to on-board data buses and other monitoring systems, is designed to extract train-rail operational parameters, modify the control variables and determine the required time for action to prevent the potential derailment accident, if a high probability of derailment exists using HIs and limited COPs. During operation, the on-board system extracts, processes, and feeds collected data and information to the on-board LSTM neural network for deriving time-based mean and oscillatory parameters of COPs. Using on-board lookup tables, the on-board computation with low fidelity models will perform its computation and continue its iterations with real-time data until any calculated or operational COPs exceed the limit COPs. When this condition is first reached, the associate sub-system is identified and control parameters (e.g., speed, braking movements) are calculated. Modified control parameters are then fed into the notional adaptive PTC scheme, which can provide the platform and operator options to manually or autonomously execute the train control authority accordingly to prevent a potential derailment accident, Figures 1 and 3.

During the operation of the trains, at a designated train operation center equipped with the high performing computing facility, the high fidelity-coupled multi-body wheel-rail interaction physics-based model and neural network with additional hidden-layers are also running for providing the model calibration and increasing the computational accuracy in real time. Solutions are then updated into lookup parameter tables. These tables, together with other relevant information including the train-rail diagnostics and prognostics data, are automatically uploaded onto the on-board system at the end of service via on-board wireless gateway transmitter and receiver. However, if anomalies are detected during operation, a set of relevant data and information for a specific faulted sub-system or low (or high) fidelity solutions can be immediately transmitted to or received from the local train operation center and enterprise server using available wireless capabilities such as Global System for Mobile Communication (GSM) and Global Position System (GPS).

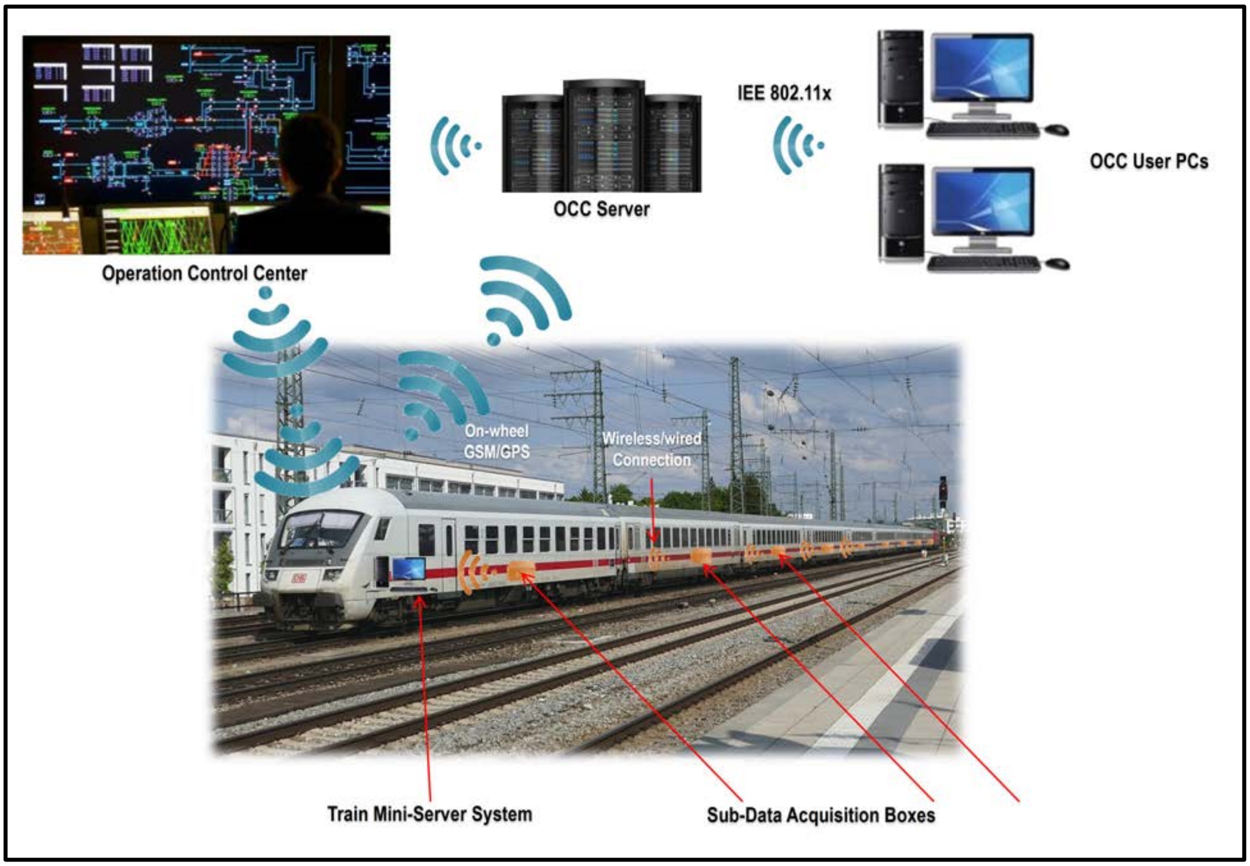

Figure 5: Machine Learning-Based Health State Awareness Scheme for Train-Rail System

2.5 STATE-OF-THE-ART MICROPROCESSOR AND HARDWARE CONFIGURATION

Numerous tech giants (e.g., IBM, INTEL, and Google) have been working on very ambitious initiatives to create processor cores, which can mimic human neurons and synapses, to accommodate the demands for high-throughput CPU and Graphic Processing Unit (GPU) designed for neural engines, machine learning algorithm, and other AI tasks. These CPUs and GPUs (e.g., for use in state-of-the-art computers, which go beyond the standard John von Newman computer architecture paradigm), are designed for the real-time power efficiency, highly complicated neural networks, and at massive scales. These microprocessors take into account of localized information processing architecture for the real-time computation communication and extremely efficient data-storage and data-crunching tasks.

While waiting for the high-end CPU/GPU prices to become less expensive, low-hanging-fruits of the powerful and highly parallel CPUs/GPUs can be configured to enable the development of the envisioned “iDaC” system for the machine learning tasks. A general outline of the core microprocessors and functionalities is provided for developing a working prototype system to be installed on a rolling stock for extensive testing and demonstration. The configuration is envisioned to allow both simplified high level programing and hardware-in-the-loop taking advance of the enhanced field-programmable field array (FPGA) for parallel processing.

An on-wheel machine learning-based computer-server system is envisioned to include: (1) a train mini-server system (TMSS) and (2) sub-data acquisition boxes (SDAB). This system, with appropriate interfaces, can be connected to other monitoring systems installed on the existing trains to process incoming data for real-time health monitoring. The envisioned iDaC system assumes data are coming from sensors, which have already been installed on all bogies in trains. In the envisioned system, the number of SDABs is proportional to the number of the train bogies. These boxes will collect data and, using the data communication scheme (e.g., consistent with the International Union of Railway (UIC) 568 standard for transmitting a variety of data between TMSS and SDABs), send it to TMSS for processing and analysis. Each SDAB, which is a standalone unit with the industry standard communication, is capable of converting the train signal information into a data stream sent to TMSS via wiring or wireless communication scheme. Figure 5 illustrates the envisioned machine learning-based health state awareness scheme for the train-rail system.

The TMSS configuration is a mini-server integrated with a microprocessor and generic board configuration to include a COTS dual-processor 32 bit (or at least 16 bit) architecture, in which each processor (e.g., one for the coupled multi-body wheel-rail interaction physics-based modeling and a second one for the AI machine learning tasks) resides on its own board. Communications between processors are done via data transmission interfaces or buses, where APIs will facilitate the assignment of software tasks to the appropriate processors. The modeling microprocessor board handles the tasks configured with FPGA for synchronizing speed controls with the notional adaptive PTC. The AI microprocessor board handles machine learning and health monitoring tasks. Since the AI board is used for complicated tasks, it is configured with much higher clock speed and large memory system and equipped with the universal asynchronous receivers and transmitters (UARTs) for communications with the Operation Control Center (OCC).

TMMS provides: (1) communication links between OCC, server, and users, (2) perform the AI machine learning diagnostics and prognostics tasks, (3) modify controls variables and determine the required time for action, and communicate the results to notional adaptive PTC for execution. TMMS can be installed in an isolated section in the operator section.

As the new central processing unit architectures, designed for AI applications, and reliable open source software are tremendously evolved, the cost for the envisioned machine learning-based health state awareness configuration for the train-rail system can be significantly affordable, easily upgradable, and flexibly scalable.

3.0 CONCLUSION

The automated and intelligent health state awareness and adaptive controls railway system cannot be effectively designed and implemented without the integration of the artificial intelligence-inspired machine learning technology. When the identified influential operating factors combined with or integrated into the appropriate physics-models, automation strategies can be potentially developed to reduce the maintenance frequency to minimum or zero within a desirable maintenance-free interval. For several decades, the research community has developed the rule and statistical learning-based AI technologies. However, there are challenges, particularly in the machine learning and training where issues with sparse data exist. Even when a vast amount of data is available, noisy data can pose a significant problem. Additionally, a realistic complex operation of the railway system can introduce an unstructured environment and hide many critical system characteristics, which might not be able to uncover unless physics-based models are integrated into an AI-inspired machine learning framework. In addition to the prior knowledge, the envisioned intelligent diagnostics and controls framework introduces the next-generation AI capabilities that can integrate the physics-based models and acquire the essential knowledge from many hidden states of complex operations and unstructured domains [11].

4.0 ACKNOWLEDGEMENTS

The author greatly appreciates the support from the Texas Tech University – Office of the Vice President for Research. Additionally, the author expresses a sincere gratitude for the encouragement from Professor Fu-kuo Chang of Stanford University and support from the Committee of the 2nd International Workshop on Structural Health Monitoring for Railway System.

5.0 REFERENCES

[1] Saat, R., Barkan, C., and Liu, X., "Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates", Transportation research record: Journal of the Transportation Research Board, No 2289, Transportation Research Board of the National Academies, Washington, D.C. 2012, pp 154-163.

[2] USDOT Small Business Innovation Research (SBIR) Program, Phase I Program Solicitation, '18-FR2: AI Tools for Relating Various Sensor Data to Adverse Vehicle/Track Interaction", Issue Date: January 17, 2018.

[3] Le, D. and et al., ""Fatigue-Free" Platforms: Vision for Army Future Rotorcraft", American Helicopter Society (AHS) 70th Annual Forum Proceedings, Montreal, Canada, May 20-22, 2014.

[4] https://kafka.apache.org/intro.html

[5] https://storm.apache.org/releases/current/Tutorial.html

[6] Lei, X., Wu, S, & Zhang, Bin, "Dynamic Analysis of the High Speed Train and Slab Track Nonlinear Coupling System with Cross Iteration Algorithm", Journal of Nonlinear Dynamics, Volume 2016, Article ID 8356160.

[7] https://deeplearning4j.org/lstm.html#long

[8] Cho., K. and et al., "Learning Phrase representations using RNN Encoder-Decoder for Statistical Machine Translation", Cornell University Library.

[9] Malhotra, P. and et al., "Multi-Sensor prognostics Using an Unsupervised Health Index Based on LTSM Encoder-decoder, arXiv: 1608.06154v1 [cs.LG] 22 Aug 2016.

[10] Bechhoefer, Eric, Augustin, M., Kingsley, M., "Architecture for a Light Helicopter HUMS", AHS 68th Annual Forum, Fort Worth, Texas, May 1-3, 2012.

[11] Le, D., "Fabrics of Artificial Intelligence-informed Technology for Holistic Sustainment (FAITHS)", Presented to Defense Advanced Research Project Agency (DARPA), Arlington, VA, March 24, 2017