EDWIN PIVA DUARTE

A MACHINE LEARNING-BASED DIGITAL TWIN MODEL FOR PRESSURE PREDICTION IN THE FUEL INJECTION SYSTEM

CURITIBA

2024

EDWIN PIVA DUARTE

A MACHINE LEARNING-BASED DIGITAL TWIN MODEL FOR PRESSURE PREDICTION IN THE FUEL INJECTION SYSTEM

Thesis presented to the Graduate Program in Computer Science – PPGIa of Pontificia Universidade Catolica do Parana - PUCPR in partial fulfillment of the requirement for obtaining a degree of Master in computer Science.

Area of concentration: Computer Science

Advisor: Eduardo Kugler Viegas
Co-advisor: Altair Olivo Santin

CURITIBA

2024

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Edilene de Oliveira dos Santos CRB 9 / 1636

Duarte, Edwin Piva
D812m A machine learning-based digital twin model for pressure prediction in the
2024 fuel injection system / Edwin Piva Duarte ; Advisor: Eduardo Kugler Viegas ;
Co-advisor: Altair Olivo Santin. -- 2024
62 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná,
Curitiba, 2024
Bibliografia: f. 58-62

1. Informática. 2. Aprendizagem do computador. 3. Automóveis – Motores - Sistemas de injeção eletrônica de combustível. I. Viegas, Eduardo Kugler. II. Santin, Altair Olivo. III. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática. IV. Título

CDD 20. ed. – 004

Pontifícia Universidade Católica do Paraná
Escola Politécnica
Programa de Pós-Graduação em Informática

Curitiba, 23 de outubro de 2024.

87-2024

DECLARAÇÃO

Declaro para os devidos fins, que **Edwin Piva Duarte** defendeu a dissertação de Mestrado intitulada "**A MACHINE LEARNING-BASED DIGITAL TWIN MODEL FOR PRESSURE PREDICTION IN THE FUEL INJECTION SYSTEM**", na área de concentração Ciência da Computação no dia 26 de setembro de 2024, no qual foi aprovado.

Declaro ainda, que foram feitas todas as alterações solicitadas pela Banca Examinadora, cumprindo todas as normas de formatação definidas pelo Programa.

Por ser verdade firmo a presente declaração.

gov.br
Documento assinado digitalmente
EMERSON CABRERA PARAISO
Data: 24/10/2024 10:17:35-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Emerson Cabrera Paraiso
Coordenador do Programa de Pós-Graduação em Informática

Abstract

Over the last years, the engine calibration task has been mostly conducted based on the engineers' knowledge. However, due to the increased complexity of modern engines, finding their most suitable configuration became an unfeasible and costly task, causing engines to be produced with inadequate calibration settings, decreasing the lifespan of their components while also degrading their efficiency. In light of this, this work proposes a machine learning digital twin model for engine pressure prediction, split into two strategies. First, we extract statistical engine features based on a predefined time window to depict the engine behavior over time. Second, a digital twin implemented through a machine learning model is used for the prediction of future engine pressure levels. As a result, the predicted values can be used to assist the engine common rail system module to avoid undesired engine states. Experiments performed in a new dataset extracted from real diesel-based engines, composed of 208 features and over 1.3 million instances have shown the proposal's feasibility. The proposed scheme is able to predict in an advance of 0.1 seconds the engine pressure levels with only 0.057 of RMSE. In addition, it increases its error rate by only 10.6% if a 0.5 second of time advance is needed.

Keywords: Digital Twin, Engine, Machine Learning.

Resumo

Nos últimos anos, a tarefa de calibração do motor foi realizada principalmente com base no conhecimento dos engenheiros. Porém, devido ao aumento da complexidade dos motores modernos, encontrar sua configuração mais adequada tornou-se uma tarefa inviável e onerosa, fazendo com que motores fossem produzidos com configurações de calibração inadequadas, diminuindo a vida útil de seus componentes e também degradando sua eficiência. Diante disso, este trabalho propõe um modelo de gêmeo digital de aprendizado de máquina para previsão de pressão do motor, dividido em duas estratégias. Primeiro, extraímos os recursos estatísticos do mecanismo com base em uma janela de tempo predefinida para representar o comportamento do mecanismo ao longo do tempo. Em segundo lugar, um gêmeo digital implementado por meio de um modelo de aprendizado de máquina é usado para prever os níveis futuros de pressão do motor. Como resultado, os valores previstos podem ser usados para auxiliar o módulo do sistema common rail do motor a evitar estados indesejados do motor. Experimentos realizados em um novo conjunto de dados extraído de motores diesel reais, composto por 208 features e mais de 1,3 milhão de instâncias, mostraram a viabilidade da proposta. O esquema proposto é capaz de prever com antecedência de 0,1 segundos os níveis de pressão do motor com apenas 0,057 de RMSE. Além disso, aumenta sua taxa de erro em apenas 10,6% se for necessário um avanço de 0,5 segundo no tempo.

Palavras-chave: *Gemeos digitais; Motores, Aprendizagem de Máquina.*

Illustrations

Tables

Abbreviations

AFRL	*Air Force Research Laboratory*
AM	*Additive manufacturing*
ASAM	*Association for the Standardization of Automation and Measurements*
CAN	*Controller area network*
CFD	*Computational fluid dynamics*
COV	*Coefficient of variation*
CRS	*Common Rail System*
DT	*Digital Twin*
DTDED	*Digital- Twin Diesel Engine Dataset*
ECU	*Electronic Control Unit*
EGR	*Exhaust gas recirculation*
hPa	*Hectopascal*
IMEP	*Indicated mean effective pressure*
IoT	*Internet of things*
Km/h	*Kilometers per hour*
MAE	*Mean absolute Error*
MBD	*Model-based definition*
mV	*millivolts*
ML	*Machine learning*
NASA	*National Aeronautics and Space Administration*
OEMs	*Original Equipment Manufacturers*
Pa	*Pascal*
PCA	*Principal Component Analysis*
PID	*Proportional–integral–derivative*
PLM	*Product life cycle management*
R&D	*Research & Development*
RPM	*Rotation per minute*
R2	*R Square*
RMSE	*Root-mean-square deviation*

List of Mathematical Notations

K_P	Proportial gain
a	Factor in avoiding K_{pcrit}
K_{pcrit}	Proportional gain value that excites the system at a resonant frequency
K_i	Integral gain
b	Factor in maintaining correlation ship between gains
f_{crit}	Frequency equal or close to natural frequency
K_d	Derivative gain
c	Factor in maintaining correlation-ship between gains
P_{real}	Real fuel injection system pressure (hPa)
P_{set}	Desired fuel injection system pressure (hPa)

Index

Chapter 1

Introduction

1.1. Context

In the last decade, companies of all sizes, all around the globe are facing ever more fast-paced, uncertain, and complex boundary conditions. A driver of this phenomenon is the growing digitization forcing companies to develop more cost and time efficiently. On the other hand, digital or virtual engineering also enables companies to cope with these challenges. In the course of this trend, a theory called Digital Twin has developed over the last two decades. The term describes the virtual representation of a physical system. In the beginning, Digital Twins were merely descriptive, but as computational and information and communication technologies evolved, it became possible to establish a bidirectional coupling between the digital and the physical system. (Grieves, 2016)

Digital twin technology refers to the creation of a virtual representation or a digital counterpart of a physical object, process, system, or entity. It involves capturing real-time data from sensors, devices, and other sources to simulate the behavior, characteristics, and performance of the physical counterpart. (Singh, et al., 2021)

Digital twins are dynamic models that mirror their physical counterparts and provide a virtual environment for monitoring, analysis, and optimization. A digital twin comprises several components, including data acquisition and integration systems, data analytics and visualization tools, and simulation capabilities. The data acquisition component collects real-time data from sensors, Internet of Things (IoT) devices, and other sources. The integration system ensures that the data is integrated and transformed into a usable format. Data analytics and visualization tools enable the interpretation and analysis of the collected data. Finally, simulation capabilities allow users to simulate different scenarios and test changes in the virtual environment. (Singh, et al., 2022)

A digital twin aims to accurately reflect the physical object properties, as represented by their sensors' values. The values depict a series of the physical object conditions, which are then used to create the Digital Twin (DT). The DT can then be used to run simulations, investigate performance issues, and generate possible improvements and insights within the digital domain, which can then be applied back to the original physical object (Schluse, 2018). Several approaches have been proposed to create digital twin, wherein authors typically resort to machine learning techniques (Horchulhack, 2022).

Machine learning algorithms can be integrated into digital twins to enhance their capabilities. By leveraging machine learning techniques, digital twins can analyze large volumes of data, detect patterns, and make accurate predictions. Machine learning algorithms can also be used to automate processes, optimize performance, and enable autonomous decision-making within digital twins.

Machine learning algorithms within digital twins can generate predictive analytics to forecast future behavior and optimize performance. By training models on historical data, machine learning can identify patterns and correlations, enabling accurate predictions. Machine learning algorithms can be applied within digital twins to detect anomalies and diagnose faults. (Goodwin, 2022) This helps in early detection of faults, enabling timely interventions and reducing the risk of equipment failure or system breakdowns.

Digital twins integrated with machine learning can enable adaptive control and optimization. By continuously analyzing real-time data, machine learning algorithms can adjust control parameters and optimize system performance in real-time. This adaptive approach enhances efficiency, reduces energy consumption, and improves overall operational performance. In summary, digital twin technology offers immense potential across various industries. Its ability to create virtual replicas of physical objects, systems, and processes enables real-time monitoring, simulation, and optimization. By leveraging digital twins with machine learning, organizations can further enhance their capabilities, gain predictive insights, and achieve higher levels of performance, efficiency, and innovation. (Goodwin, 2022)

1.2. Motivation

Developing new vehicle engines, especially those based on diesel fuel, is a challenging task that demands manufacturers to meet government legislation while ensuring engine efficiency (Knecht, 2008). To fulfill such a task, manufacturers must properly calibrate the fuel injection system, which is directly related to the exhaust gas system and the Common Rail System (CRS). In general terms, the common rail system calibration is based on a proportional–integral–derivative (PID) controller embedded in the Electronic Control Unit (ECU). The controller's goal is to fine-tune the injection system's real pressure based on the ECU's required pressure signal. Engine manufacturers have been performing the calibration task based solely on engineers' expertise following a previously settled guideline (Lygoe, 2010). Therefore, the calibration is only mutable during the development phase, and several quality levels must be met before the engine production can start.

Original Equipment Manufacturers (OEMs) tackle multiple challenges, particularly during the engineering phase, where the fine-tuning of engine calibration becomes very important. Engine calibration, the process of optimizing engine parameters to achieve desired performance, efficiency, and emissions targets, is a delicate balancing act, necessitating a deep understanding of engine dynamics, combustion phenomena, and regulatory constraints. One of the biggest challenges confronting OEMs in this industry is achieving optimal trade-offs between conflicting objectives, such as power output versus fuel consumption and emissions reduction (Meli, 2024). Discovering an equilibrium between these opposing demands requires a complex calibration process, characterized by iterative adjustments to fuel injection timing, air-fuel ratio, turbocharger boost pressure, and exhaust gas recirculation (EGR) rates. Additionally, meeting emission standards, coupled with the dynamic nature of market requirements, aggravates the complexity of calibration tasks, leaving OEMs to adopt innovative methodologies and robust optimization techniques to handle this task effectively.

Moreover, the different operating conditions makes it a challenging task to calibrate due to different ambient temperatures, altitudes, and load profiles. Diesel engines are found in a wide range of environments starting from the extremely cold Arctic or Antarctic regions all the way down to hot desert terrains that pose several challenges regarding engine performance and emissions control. The calibrated model must incorporate robust control strategies that tune the engine parameters as they work within real time constraints for adjustments to reach the best performance and control the emissions (Laubichler, 2024). Transient operating conditions (such as rapid acceleration/deceleration) bring additional complexities calling for quick responses from engine management system. These concerns can only be addressed through approaches that combine advanced sensors, predictive algorithms, and model-based controls to make these calibration systems more flexible and responsive.

The search for innovation in the competitive automotive landscape highlights the pressure on OEMs to deliver cutting-edge diesel engines that outperform rivals in terms of performance, efficiency, and reliability (Sandberg, 2024). In an era characterized by rapid technological evolution, staying ahead of your competitors by continuous refinement of calibration methodologies, leveraging advancements in computational fluid dynamics, machine learning, and virtual prototyping is needed to survive. However, this quest for innovation must be tempered with careful validation and verification processes to ensure robustness and compliance with regulatory standards.

Balancing the constraints of innovation and validation represents a big challenge, where OEMs must be careful to meet the needs of customers and regulatory authorities. The calibration of diesel engines emerges not merely as a technical challenge but as a strategic necessity, shaping the competitive environment and defining the success trajectory of OEMs in the global automotive arena. (Nicoleta, 2011)

In the scope of diesel engine development, the beginning of the common rail fuel injection system brought new opportunities, offering enhanced precision, flexibility, and control over fuel delivery. However, this technological advancement has brought a new set of challenges, and one of the main ones being the management of overshoots and undershoots in fuel pressure regulation, particularly during transient operating conditions. Overshoots and undershoots refer to deviations from the desired fuel pressure setpoint, occurring as a result of dynamic changes in engine load, speed, and operating conditions. (Lu, 2024) Effectively addressing these phenomena requires a multifaceted approach, encompassing advanced control algorithms, robust hydraulic components, and sophisticated sensor technologies.

As each system calibration comes out uniquely, finding possible failures and their root causes has become increasingly challenging considering the rising number of projects and their complexity (Xi, 2018). As an example of failure, a not adequately tuned controller may generate overshoots and undershoots that the CRS might face during its usage. Overshoots occur when the real pressure of the system exceeds the required pressure limits, reaching higher than desired levels. Undershoots lowers the system pressure beyond its intended operation level. Such situations may significantly decrease the engine lifespan while degrading its efficiency. In such a context, over the last years, several works have been proposed to improve the industry engine development process, wherein approaches based on digital twin have yielded promising results (Bhatti, 2021).

One of the primary challenges in mitigating overshoots and undershoots lies in the fundamentally nonlinear and time-varying nature of the common rail fuel injection system. Traditional control strategies, such as proportional-integral-derivative controllers, often struggle to anticipate and compensate for rapid fluctuations in fuel demand, leading to transient pressure deviations. The presence of hydraulic dynamics, including pressure waves and fluid inertia, intensifies the complexity of the control problem, necessitating the development of model-based control techniques capable of capturing system dynamics with high fidelity (Soltanalizadeh, 2024). Likewise, the interaction between various components within the common rail system, such as injectors, high-pressure pumps, and pressure regulators, further complicates the control task, requiring a general approach to system modeling and control synthesis.

Another critical challenge stems from the stringent performance requirements imposed by emissions regulations and customer expectations. Overshoots and undershoots in fuel pressure can have adverse effects on engine combustion dynamics, leading to increased emissions, reduced efficiency, and compromised drivability. Achieving compliance with emissions standards while simultaneously delivering optimal engine performance necessitates precise control over fuel delivery dynamics, necessitating the development of adaptive control strategies capable of dynamically adjusting control parameters in response to changing operating conditions (Deffo, 2024). Furthermore, the integration of predictive algorithms, leveraging machine learning and data-driven modeling techniques, can enhance the anticipatory capabilities of the control system, enabling proactive mitigation of transient pressure deviations before they manifest in engine output.

A behavioral model can be built by evaluating a training dataset composed of vast amounts of the physical object collected sensor values. As a result, the built model, acting as the DT, will be able to portray the data used during the training phase. Unfortunately, the building of a realistic DT training dataset is a challenging task. A realistic DT must collect training data under a variety of conditions. This should include faulty and normal events often not easily achieved during the engine development. The DT can often only portray the relationship between the collected sensor values even if a realistic training dataset is available. Therefore, the way they can be used for improvement purposes is neglected, for example, to improve system calibration in an engine fuel injection.

1.3. Goals

This document proposes a CRS DT model based on machine learning (ML) techniques to assist the engine calibration task, implemented through two strategies. First, we extract statistical features based on a predefined time window of the collected engine sensor values. The insight of such an approach is that historical statistical features can be used to represent the engine behavior in detecting future engine failures. Second, the extracted feature values are used as input by a machine learning model, which acts as the engine DT. The model is used to predict pressure levels in the fuel injection system. The main insight of such an approach is to use a DT to assist the engine calibration reliability while also predicting the system behavior and engine failures so that counteractions can be performed accordingly.

To reach this main goal, the following specific goals must be met.

A. Develop a new dataset

A new dataset must be created to enable the digital twin model to be applied. This data set will contain the historical statistical features of the diesel engine. The dataset must be composed of huge amounts of samples from the environment, given that the machine learning model will be built accordingly. As it is described in this document, some data mining and pre-processing techniques must be applied to enable the development and evaluation of the digital twin machine learning algorithms.

B. Develop digital twin machine learning algorithm

A new digital twin machine learning algorithm will be built using ML techniques combined with a digital twin framework and strategy.

C. Develop a model prototype

Based on the new dataset with real data, and the algorithm developed, a model prototype will be proposed to verify and meet the main goal of this work.

D. Evaluation of machine learning algorithms

The digital twin machine learning algorithms will bring results based on the new data set created. These results are evaluated based on selected metrics to validate the proposal, showing how the undershoots and overshoots can be anticipated in the development of diesel engines.

The evaluation aims at answering the following research questions (RQ):

- (RQ1) How does our proposed DT model work for predicting pressure levels in the fuel injection system?
- (RQ2) What is the prediction performance of our model for a longer prediction time for the pressure levels?

1.4. Contributions

The main contributions of this document can be described in two main categories, a new digital twin dataset and a new digital twin model.

A. Digital twin dataset

A new digital twin dataset built through the collection of 208 sensor values from real diesel-based engines. It is composed by over 1.3 million of samples, corresponding to 10 minutes of data collection, including 57 and 52 thousand undershoot and overshoot failures, respectively. This data was collected in a diesel engine vehicle still in validation phase, where failures are still expected.

B. Digital twin model

A new digital twin model based on machine learning techniques for the prediction of pressure levels in the fuel injection system. The proposed scheme can predict in advance of 0.1 seconds the pressure levels of the fuel injection system with only 0.057 of Root-mean-square deviation (RMSE).

1.5. Publications

This study was published in IECON 2022 (Qualis A2), the 48th Annual Conference of the IEEE Industrial Electronics Society (IES), focusing on contemporary industry topics ranging from electronics, controls, manufacturing, to communications and computational intelligence. The published paper named “A Machine Learning Digital Twin Model for Engine Pressure Prediction” by the authors Edwin P. Duarte, Eduardo K. Viegas and Altair O. Santin was presented in Brussels.

1.6. Document structure

This document is organized as follow:

- Chapter I brings the introduction to the proposal.
- Chapter II describes ML-based digital twin, bringing the background research on digital twin technologies, use cases, when machine learning is applied with this new technology. The pipeline required to develop a digital twin model, and the challenges when developing such a solution.
- Chapter III describes related works on engine fault detection, using different techniques on how to face the main challenges.
- Chapter IV presents our new dataset for engine fault detection and a preliminary evaluation.
- Chapter V presents our proposal, describing the work done related to the data mining and the build of the digital twin machine learning model.
- Chapter VII evaluates the digital twin machine learning performance based on the real dataset.

- Chapter VIII concludes the work with our outlook and proposed future works.

Chapter 2

Background

This chapter will bring the main concepts related to digital twin development and examples of use cases when a digital twin model brings benefits if correctly applied. The development of machine learning algorithms, with specific data-sets pre-processing and data mining techniques.

2.1. Digital Twin

The concept of DT was first proposed by Grieves at the University of Michigan in 2003. Grieves stated that product life cycle management (PLM) is intended to be the informational equivalent of being in physical possession and examining an item; this was considered the prototype of the DT concept. In the following years, Grieves updated his theory and named the concept a “mirrored spaced model” and an “information mirroring model”. (Grieves, 2016)

In 2010, the National Aeronautics and Space Administration (NASA) introduced the concept of DT in the space technology roadmap (Shafto, et al., 2010), intending to use DT to implement comprehensive diagnosis and prediction functions for flight systems to ensure continuous and safe operation during their service life.

The Air Force Research Laboratory (AFRL) introduced a conceptual model, in 2011, that used DT technology to predict the life of aircraft structures and gradually extended it to the airframe condition assessment study. They also combined historical flight monitoring data to virtual flight to assess the maximum allowable load while ensuring airworthiness and safety, thereby reducing the life cycle maintenance, and increasing aircraft availability. Meanwhile, the US Department of Defense designed a digital thread (a physical based model instance), which became the basis for cross-domain data exchange.

It is possible to find many different explanations and definitions for digital twin in the literature. For example, in 2012, Glaessgen and Stargel first defined DT as an integrated Multiphysics, multiscale and a probabilistic simulation (Glaessgen, 2012). This simulation used physical models, sensor updates, and fleet history, to mirror its corresponding flying twin's life. With the recent improvement of sensing, software, and hardware technology, combined with computing performance improvements, the DT concept has been further developed, especially in the real-time operation monitoring of products and equipment.

Another example is the development of machines that involves complex system engineering, development requirements, system composition, product technology, manufacturing processes, test and maintenance, project management, working environment, and other issues. As a possible way to realize the interactive integration of the physical world and the digital world, DT is gradually applied to all aspects of the product life cycle, including product design, industrial production, and manufacturing services (Xie, et al., 2021). Use of DT for improving R&D quality, manufacturing, production efficiency, and predictive maintenance of equipment is of great significance.

The DT application framework for the complex system closed-loop life cycle is depicted in Figure 1. It is important to note that the DTs have different forms at different stages of the whole life cycle. Specifically, in the design and test stage where there is no physical entity. In this case, the requirement for the DT is the user's requirements. By inputting the quantitative user requirements into the DT and modifying its model, the system design's reliability can be predicted. In the manufacturing/assembly and operation/service stages, the physical asset corresponds to the DT model (Xie, et al., 2021). The physical system measurement parameters are transmitted to the DT in real-time to realize the high fusion of virtual and real entities. In the scrapping/recycling stage, DT can be extended to the next cycle's development process, forming a closed-loop life cycle, even though the physical entity does not exist.

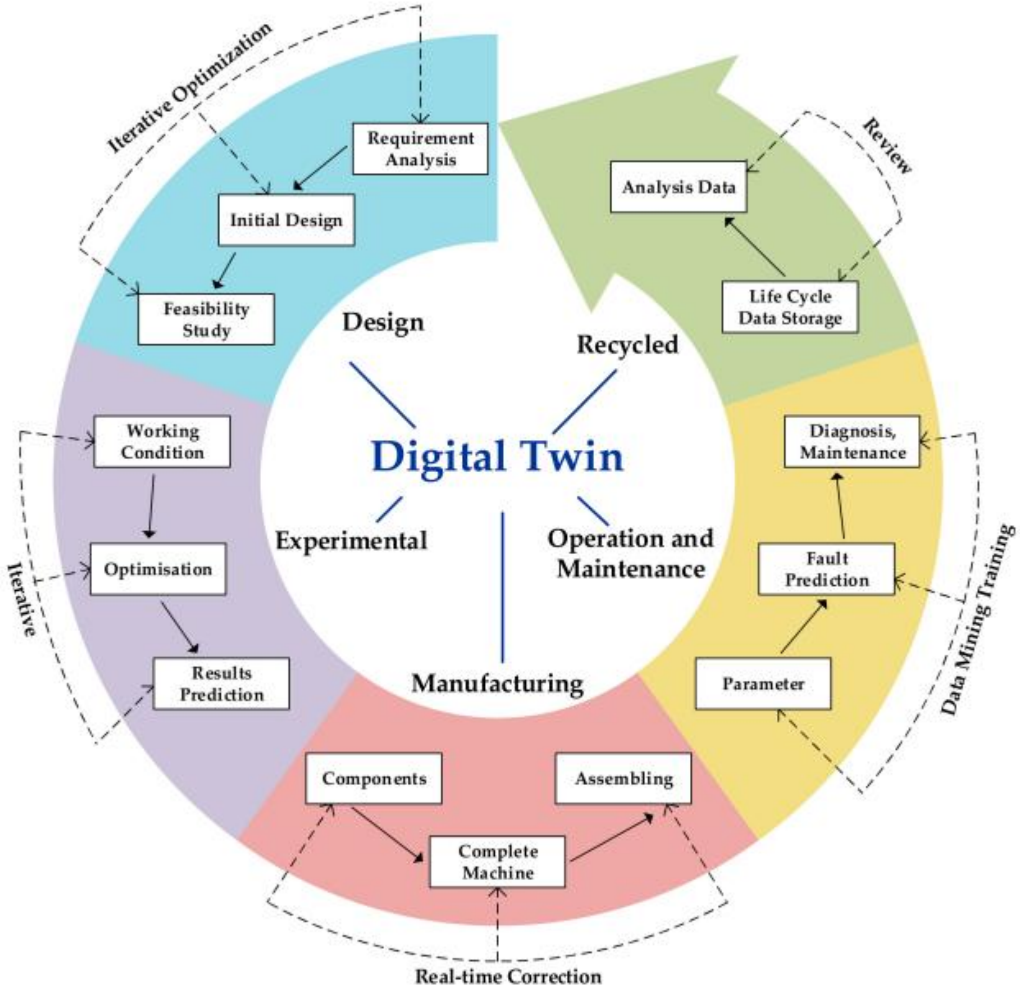

Figure 1 - A general LCM framework with digital twin - image extracted from (Xie, et al., 2021)

A digital twin (DT) aims to reproduce the behavior of a complex physical product using a probabilistic function, which is used to mirror the behavior of its corresponding twin (Anderl, 2018). Its main goal is to act as a digital copy of a physical asset. As a result, a DT can give assessments of how a system will perform under production. Thus, it can be used to identify and easily pave the way to efficiency improvements. A DT is typically utilized at the initial phases of development and design. It enables the precise reproduction of how specific systems and subsystems will perform in a set of predefined circumstances.

2.2. Digital twin use cases

Many DT technologies play an essential role in the life cycle of systems design, manufacturing, operation, and maintenance. In the product design stage, the model-based definition (MBD) technology realizes the efficient expression of product data, and the lightweight model technology optimizes the model's storage structure. The simulation and optimization technology makes the product DT model closer to the physical product's functions and characteristics. In the manufacturing and assembly stage, DT uses multi-level interconnection such as industrial Internet, IoT, and sensors, and collaborates with information technologies such as artificial intelligence, machine learning, data mining, and high-performance computing. These technologies play an essential role in multisource structure data collection, data integration display, product production supervision, quality management, intelligent analysis, and decision-making in complex dynamic spaces. In the operation and maintenance phase, DT comprehensively utilizes sensor technology, traceability technology, simulation technology, and IoT technology to support status tracking and monitoring, early fault warning, life prediction, and positioning analysis. The conceptual diagram of a wind turbine realizing DT is shown in Figure 2

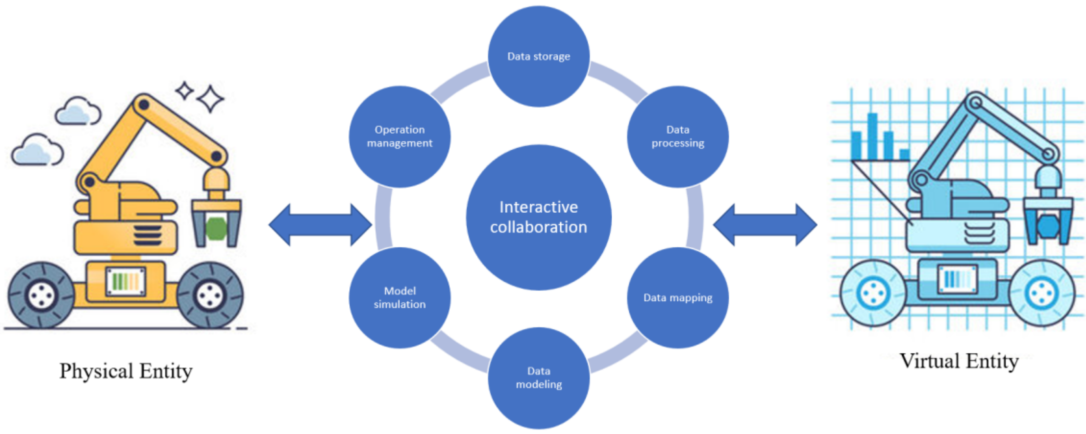

Figure 2 - Conceptual diagram of digital twin

Aerospace/aeronautics fields were the pioneer areas where DT was explored first by NASA and by the U.S. Air Force. The main applications of DTs in this industry include optimizing the performance and reliability of the space vehicle/aircraft, predicting, and resolving maintenance issues, and making the missions safer for the crew. The main application of DT in this industry started with the intention to optimize the performance and reliability of the space vehicle/aircraft. According to the (Shafto, et al., 2010), the four applications of DT for them were:

- Simulating the flight before the launch of the actual vehicle to maximize the mission success.
- Continuously mirroring the actual flight and updating the conditions such as actual load, temperature, and other environmental factors to predict future scenarios.
- Diagnosing damage caused to the vehicle.
- Providing a platform to study the effects of modified parameters that were not considered during the design phase.

According to Singh, although the development of DT started from the Aerospace industry, the industry which is exploring the technology the most is the manufacturing industry. (Singh, et al., 2022). Any manufactured product goes through four main phases throughout its life cycle: design, manufacture, operation, and disposal (Figure 3). Smart manufacturers can leverage DTs in all four phases of the product

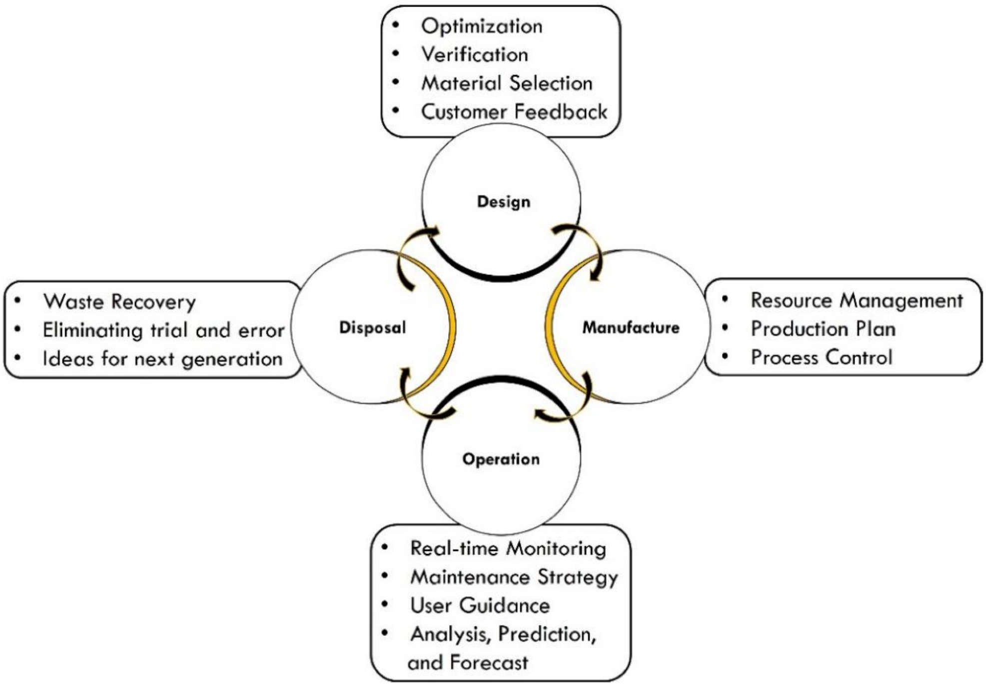

Figure 3 - DT's applications throughout a product's lifecycle - image extracted from (Singh, et al., 2022)

During the design phase, Digital Twins enable designers to virtually validate their product concepts, allowing them to explore various iterations and select the most effective design. By leveraging real-time data from prior product models, designers gain valuable insights into which features resonate with consumers and which aspects may require enhancements. This streamlines and accelerates the design improvement process. A case in point is Maserati, which utilized DT to enhance the aerodynamics of automobile bodies through virtual wind tunnel simulations, a traditionally complex and costly approach. They also improved the interior acoustics of their vehicles by employing data collected from a microphone-equipped dummy during prototype testing. (Shao & Helu, 2020)

The subsequent stage in manufacturing involves transforming raw materials into finished products. At this point, a Digital Twin can play a vital role in managing resources, planning production, and overseeing process controls. (Tao, 2018)

Once the product is sold and in use, manufacturers can access real-time operational data through its corresponding Digital Twin, enabling them to devise maintenance strategies effectively. Fei Tao et al. noted that a Digital Twin can offer nine types of product services, including real-time status monitoring, energy consumption analysis and forecasts, user management and behavior analysis, operational guidance for users, intelligent optimization and updates, failure analysis and prediction, maintenance strategy development, virtual maintenance, and virtual operations. (Tao, 2018)

The final stage of a product's life, often overlooked, pertains to disposal. Consequently, valuable insights that could inform future product or system developments are frequently lost upon the product's retirement. Xi Vincent Wang and Lihui Wang ((Liu, 2022) introduced an innovative Digital Twin-based system designed to facilitate the recovery of waste electrical and electronic equipment, supporting manufacturing and remanufacturing efforts throughout the product life cycle from design to recovery.

As already mentioned in this document, another use case for DT models is in the development of new engines, such as diesel-based engines. The development of engines is extensive, complex, and expensive especially when dealing with the critical system that requires calibration such as the common rail injection system (CRS).

A CRS layout, as shown in Figure 4, which is commonly used by traditional diesel-based engines, such as the one used by a commercial vehicle with a 2.0 liters diesel engine 4×4 model. In the CRS, the pressure generation and the fuel injection are independent. The pressurization of the fuel takes place in the Common Rail, as the high-pressure pump supplies a continuous flow of diesel to it, ensuring the fuel is under the ideal pressure and ready to be injected.

In common rail systems, a high-pressure pump stores fuel in a tank at pressures reaching and exceeding 2,000 bars (200 MPa). The term "common rail" indicates that all fuel injectors are supplied from a single fuel rail, essentially a pressure accumulator, where the fuel is maintained at high pressure. This accumulator ensures that multiple injectors receive fuel at the required high pressure. Consequently, the role of the high-pressure pump is simplified, as it only needs to maintain the desired pressure, which can be controlled either mechanically or electronically.

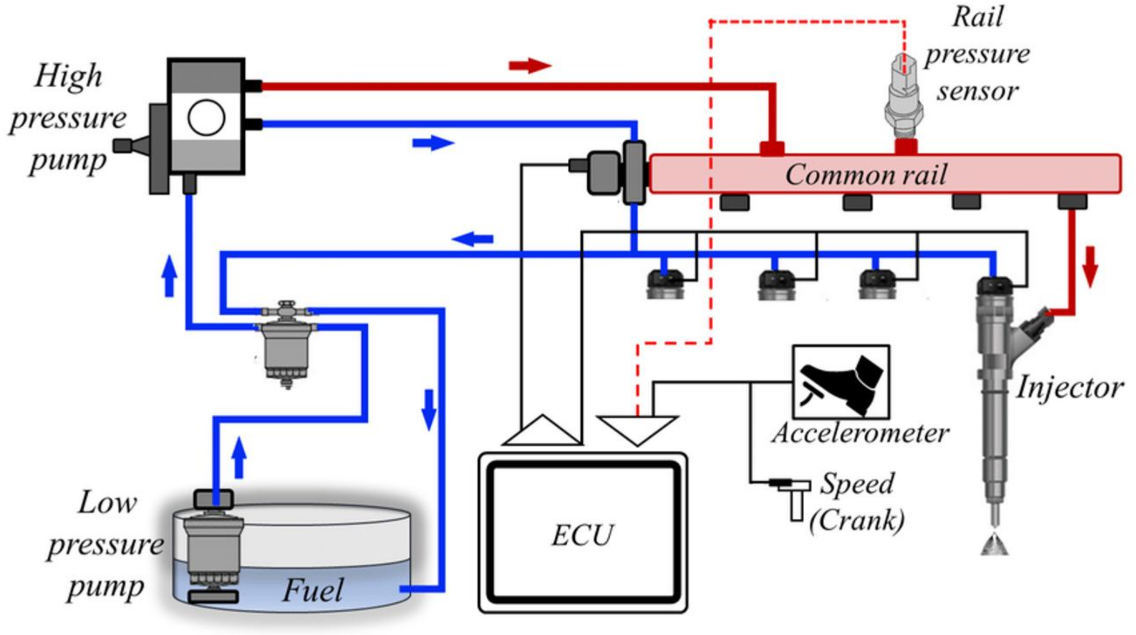

Figure 4 - Common rail injection system (CRS) layout is considered in our work. The layout depicts a CRS in a real vehicle as used in a commercial vehicle with a 2.0 liters diesel engine.

The CRS system generally consists in supplying all cylinders by the common fuel rail. The system includes measuring elements, central control unit, low pressure pump, high-pressure pump, common rail, and injectors.

To enable its proper operation, it is a must that the manufacturer ensures the appropriate calibration of the CRS, which is typically achieved by a proportional–integral–derivative (PID) controller embedded in the Electronic Control Unit (ECU). The main calibration goal is to ensure that the pressure of the injection system reflects the required pressure signal from the ECU. Therefore, CRS operation can be typically described according to three main situations:

- Normal. Expected CRS injection system state wherein the system injects the expected fuel amount as computed by the PID controller. The CRS operates at the proper pressure, resulting in no waste of fuel and no degradation of the system components' lifespan.

- Overshoot. Failure state wherein the CRS injection system over-injects fuel. Fuel injection system pressure in- creases, causing fuel waste and degradation of the engine components' lifespan.

- Undershoot. Failure state wherein the CRS injection system under-injects fuel. Fuel injection system pressure decreases beyond expected, affecting the engine reliability, and driving comfort.

Even though overshoots and undershoots are very common on diesel-based engines with CRS using the traditional PID controllers, the goal of the calibration phase is to eliminate them completely, unfortunately, due to the PID limitations, that is usually not the case. The real values of this type of miscalibration in engines are not shared by the manufacturers, as it directly impacts the lifespan of their engines and the fuel consumption of their vehicles.

2.3. Machine learning

Machine learning algorithms play a pivotal role in numerous applications, from recommendation systems to medical diagnostics, offering unprecedented opportunities for automation and decision-making. However, their development poses multifaceted challenges, spanning from data acquisition to model evaluation. Central to the process is the extraction of valuable insights from vast datasets, a task facilitated by sophisticated data mining techniques. As we embark on the journey of understanding machine learning intricacies, it becomes imperative to dissect the data mining process and unravel its significance in algorithmic development. (Carpenter, et al., 2018)

Data mining serves as the basis of machine learning endeavors, which involves a range of methods for discovering patterns, relationships, and anomalies in raw data. The key is to extract useful information from different sources that will require an intelligent choice of mining algorithms based on specific goals (Chen, 2024). Data mining can be approached through two main techniques: classification and clustering. Classification algorithms are designed to assign pre-defined labels to instances based on attributes for predictive modeling and pattern recognition purposes. Where clustering algorithms attempt at decomposing data into cohesive groups unveiling its underlying structures without knowing beforehand class labels (Kameshwaran, 2014). These approaches complement each other thus providing a complete grasp of data dynamics required to build subsequent models.

Pre-processing serves as a critical precursor to model construction, encompassing a suite of techniques aimed at enhancing data quality and compatibility with machine learning algorithms. Raw data often holds inconsistencies, missing values, and irrelevant features, posing formidable challenges to algorithmic efficacy. Supervised pre-processing techniques lead to the suggestion of missing values, normalization of feature scales, and dimensionality reduction through techniques such as Principal Component Analysis (PCA) or feature selection (Kotsiantis, 2006). Equally, unsupervised pre-processing methods, such as outlier detection and noise removal, strive to cleanse the data of spurious artifacts, raising a more robust learning environment. By navigating the pre-processing environment, practitioners can mitigate data-related impediments and unleash the full potential of machine learning algorithms.

Several approaches have been proposed for the creation of a DT in a variety of fields, including industry, medical, and even for monitoring purposes (Bulle, 2020) and (Viegas, et al., 2020). Authors generally resort to machine learning techniques, yielding promising reported results (Viegas, et al., 2020). In such a case, a machine learning model is built through a computationally expensive model training process, which evaluates the data available in a training dataset. The dataset must be composed of huge amounts of samples from the environment, given that the machine learning model will be built accordingly. As a result, the building of a realistic DT through machine learning means becomes a challenging task. This is because the data collected from the physical object sensors often cannot correctly depict the behavior of the to-be-digitalized object given that it must be monitored for a long time under a variety of environmental conditions. In such a case, the built machine learning model, acting as the DT, may present high accuracy in physical object representations. However, at the same time, it does not provide the expected level of realism to be used as a DT, e.g., to evaluate the physical object behavior under production or even pave the way to efficiency improvements.

A data set for machine learning-based digital twin model serves as the bedrock upon which accurate and robust simulations are constructed. Such a dataset summarizes a comprehensive representation of the underlying physical system, filtered from real-world observations and measurements. Including raw sensor data, historical records, and operational logs, this dataset forms the substrate from which machine learning algorithms glean insights, patterns, and relationships, facilitating the creation of virtual replicas that mirror the behavior and dynamics of their physical counterparts. (Anon., 2021)

This dataset is submitted to rigorous preprocessing and cleaning procedures to repair anomalies, address missing values, and standardize formats, ensuring its suitability for analysis and modeling. This preprocessing phase involves numerous tasks, including data cleaning, transformation, normalization, and feature engineering. Through these processes, the dataset is refined and enhanced, imbuing it with the integrity, fidelity, and relevance requisite for the development of accurate and actionable digital twin models.

There is a data set for machine learning-based digital twin model which is used to capture the detailed and diverse relationships between phenomena that occur within the physical system. Exploratory data analysis (EDA) techniques are used in order to identify the underlying patterns, relationships and trends in the dataset (Sajal, 2024). This provides important information on how the system behaves and changes over time. In addition, visualizations, summary statistics and correlation analyses can provide stakeholders with an overview of its main features so that they can be taken into account when performing subsequent modeling exercises or selecting predictors.

At last, a dataset for a machine learning based digital twin acts as an interface where practical insights and predictive power are derived from real-world experiences. Digital twin models are created through combination of real data and machine learning algorithms which mimic physical asset behaviors, enabling prescriptive decision-making capabilities, predictive maintenance actions as well as performance optimizations. Also, this potential has been enhanced by advents in areas such as data collection methodologies, pre-processing tools together with advanced analytics. Technology associated with digital twin is now potentially capable of changing industries entirely leading us into a new era of efficiency.

2.4. Challenges

The utilization of real data from diesel engine vehicles brings a new set of challenges in the development of machine learning-based digital twin models. Unlike synthetic or simulated datasets, real-world data brings complexities, uncertainties, and characteristics fundamental to physical systems, presenting unique barriers to data preprocessing, analysis, and modeling. In this chapter, we explore the challenges encountered when working with real data from diesel engine vehicles, shedding light on the difficulties of feature engineering, data quality assurance, and domain-specific knowledge integration.

The biggest challenge in using real data from diesel engine vehicles is the variety and collection of features contained in this dataset. Diesel engine systems are made up of various interconnected components each one producing different sensor data streams which monitor different operational parameters such as: temperature, pressure, fuel consumption, and exhaust emissions. Merging these types of information sources to create an input representation for machine learning entails serious difficulties that need thoughtful consideration about selection of variables as well as dimensionality reduction and representation learning methods. Also, the dataset is characterized by a dynamic nature of operating conditions; environmental factors as well as user behaviors that have unique complexities thereby demanding adaptive modeling approaches having temporal and spatial dependencies within it.

Ensuring the quality and integrity of real data from diesel engine vehicles constitutes another challenge in digital twin development. Raw sensor data may be susceptible to noise, outliers, and measurement errors stemming from sensor malfunctions, environmental interference, or data transmission issues. Additionally, missing values, incomplete records, and data drift may further complicate the preprocessing and cleaning process, necessitating robust techniques for anomaly detection, imputation, and error correction. Discrepancies between sensor readings and ground truth measurements may arise due to calibration errors or instrumentation inaccuracies, challenging the validity and reliability of the dataset. Addressing these data quality issues requires meticulous attention to detail, extensive validation procedures, and domain-specific expertise to discern genuine signals from spurious artifacts.

When dealing with actual data from vehicles with diesel engines, it is difficult to embed information from the area of specialization into the process of making models. The physical processes and thermodynamics of diesel engine system are complex and intricate due to multiple interactions among components, fluids, and combustion process. In machine learning accurate and interpretable models must be developed, which requires a deep comprehension of these foundation mechanisms as well as domain knowledge about system dynamics, performance metrics and failure modes. Nonetheless, there are challenges in expressing such knowledge as actional features or interpretability criteria for models' constraints. Therefore, this is where data scientists have to collaborate with domain experts and engineering practitioners in order to effectively link theory to practice.

Despite these challenges, the utilization of real data from diesel engine vehicles offers unparalleled opportunities for advancing the state-of-the-art in digital twin technology. By confronting complexities, uncertainties, and idiosyncrasies head-on, researchers and practitioners can leverage real-world insights to construct accurate, robust, and actionable digital twin models that empower proactive decision-making, predictive maintenance, and performance optimization in the automotive domain. As advancements in data acquisition, preprocessing, and modeling techniques continue to evolve, the potential of real data-driven digital twins to revolutionize diesel engine development and enhance operational efficiency grows exponentially, driving innovation and resilience in an increasingly dynamic and interconnected world.

In addition to the mentioned challenges of working with real data from diesel engine vehicles, the integration of a PID controller for detecting overshoots and undershoots in the common rail system introduces further complexities to digital twin development. A PID controller, a widely used feedback control mechanism, aims to regulate the fuel pressure within the common rail system by adjusting control parameters based on error signals derived from deviations between desired setpoints and actual measurements. However, tuning a PID controller to effectively mitigate overshoots and undershoots poses a non-trivial task, as it requires a deep understanding of system dynamics, response characteristics, and control loop stability.

The interaction between real data features and the PID controller presents a unique challenge in digital twin development, as the efficacy of the controller hinges upon the accuracy and relevance of input features derived from the dataset. Features such as engine load, temperature, and fuel flow rate serve as critical inputs to the PID controller, influencing its decision-making process and control actions. However, the dynamic nature of these features, coupled with uncertainties and noise inherent in real-world data, complicates the task of extracting meaningful signals and patterns that inform controller behavior. The time-based and spatial dependencies within the dataset may exhibit nonlinearities, discontinuities, or anomalies that challenge traditional control strategies, necessitating adaptive and robust control techniques capable of adapting to changing operating conditions and system dynamics.

However, additional complexities and considerations are introduced when digital twin results are employed as inputs to the system itself in order to improve calibration quality. The feedback loop between digital twin predictions and real-world control actions must be validated, calibrated, and synchronized properly to maintain consistency, stability, and convergence. This implies that any differences or inconsistencies between the outputs of a digital twin and actual responses from a system may result in inaccurate decisions for control purposes hence poor performance or even instability within the control loop. It is therefore important for the error correction, feedback integration mechanisms and model refinement process to be put up so as to fully exploit the potential of digital twin technology towards improving calibration quality along with control performance.

Chapter 3

Related works

Many researchers have proposed different approaches for studying and implementing the digital twin concept with or without machine learning, not only for diesel engines but for many other areas of the industry. This chapter highlights the related works to the proposed research, with different techniques, approaches, and proposals

As broadly discussed, digital twin concept is a famous area of research, however without many applications developed. Ghanishtha B. et al. explains the fundamentals regarding digital twin and tries to bridge the gap of individual research to provide a comprehensive review of digital twin outside of the aerospace sector (Bhatti, 2021). Aiming that, it is proposed the workflow of building and utilizing a complex digital twin with the main three stages of DT development in vehicle as archetype modeling, virtual sensors modeling and parameter update. Throughout the document many examples are giving on where this type of solution can be used with regards to vehicle development, such as:

- Predictive mobility and autonomous motion control.
- Advanced driver assistance systems.
- Vehicle health monitoring and management.
- Battery management system and intelligent charging

All these examples expect to be developed using machine learning as an enabler for digital twin technology.

Digital Twin frameworks are clearly impacting the entire product life cycle management even still being in its theoretical stage. Zheng et. al presents a systematic study about digital twin technology and application (Zheng, 2019) . The study says that Application framework of DT for product lifecycle management consists of three parts, physical space, virtual space, and information-processing layer. In the application process, the DT technology can realize the full-physical system mapping, the life-cycle dynamic modeling and the whole process real-time optimization. The bidirectional mapping and interoperability of physical space and virtual space are realized through data interaction. Intelligent decision is realized through iterative optimization and regulatory interaction between two spaces.

For instance, Airamadan A. et al. showcased the strength of machine learning models in imitating the operation of an advanced engine concept - the gasoline compression ignition at low loads (A. S. Airamadan, 2022). On this work, the authors show that machine learning model can be a useful tool in guiding the engine calibration process, however, without the use of the digital twin technology. It was showed that the machine learning models can predict seven engine performance parameters: fuel consumption, four engine-out emissions, exhaust temperature, and coefficient of variation (COV) in indicated mean effective pressure (IMEP). In this work, an experimental dataset from a study was used to development the ML model and tested in a single-cylinder engine equipped with a gasoline direct injector.

As the authors mentioned, coupling machine learning models with suitable optimization algorithms can increase the quality with less cost of the traditional engine calibration approaches. They were able to achieve good results with the four different machine learning models capturing the complex relationship between the input calibration parameters and the desired outputs.

On the other hand, M. Hinrichs used a traditional model-based approach to detect faults of a heavy-duty diesel engine based on the injected fuel (Hinrichs, 2021). The development of the study faced many challenges when handling modern’s engine control units (ECUs) with many monitoring and diagnostic functions. Many boundary conditions had to be established to enable the traditional model-based approach study to be carried out, such as securing that if the fuel is changed without adapting the ECU, a lot of faults will indicate interfering in the final result. The author was able to present new approaches to fault detection for pure diesel fuel, such as calculating the fuel mass based on the rail pressure signal, or as measuring the oxygen amount in the exhaust gas. Based on those a combustion model was developed to calculate the burnt fuel mass in combination with the intake air flow of the engine. The author presented that despite achieving satisfactory results, there are limitations in such traditional models. For example, the results can be biased considering the difficulty of applying the model in different scenarios of engine functionality, such as rail vibrations and load points on the dataset.

Similarly, di Gaeta et al. also worked with a common rail injection system technology (Gaeta, 2012). In this document, it was proposed a model-based gain scheduling approach for controlling the common-rail system for gasoline direct injection engines, aiming for best performance concerning emissions, fuel economy, drivability, and diagnostics. The proposed work made possible to vary the injection pressure over the entire engine speed range. The proposed method can be formalized as a triple-step procedure:

- A steady-state control is deduced, playing the similar role of the map-based control strategy widely used in automotive control.
- A feed-forward control is derived concerning the variation of the tracking reference.
- The previous two steps result in an explicit and affine expression of the tracking error dynamics, based on that a non-linear error feedback can be easily designed for enhancing the closed-loop performance and rearranged into a state-dependent proportional–integral–derivative (PID) controller.

Using experimental data, the authors analyzed different work conditions and with mathematical models they were able to validate and describe the pressure in the rail. On that sense, the model-based controller of the mean value pressure in the common rail device for gasoline engines was proved to be a suitable algorithm for effective implementation in commercial ECUs.

Some diagnosis studies can also be found in the literature, such as J. Zheng et al. presented the possibility of using a classification algorithm to diagnose faults of the injector used in a diesel engine common rail system (Zheng, 2020). The proposal shows the prediction of the injection volume by Gaussian Process Regression model, according to the pressure of injector pressure chamber, the fault classification is carried out by machine learning based on the time domain feature of the pressure curve when the injection volume is out of tolerance. Using a model to simulate the common rail pressure, they used features such as pressure accumulator pressure, drive signal and fuel injector flow rate. The authors achieve promising results with an average error of the predicted data of only 0,24%. If the predicted data, an SVM algorithm is used for classifying the fault according to the time-domain characteristics, with a classification accuracy as high as 93.7%. The outcome does not focus on the impact of the results on the calibration process itself, neither on any feedback towards the engine control unit.

W. Chatlatanagulchai et al. proposed a quantitative-feedback theory controller designed and applied to a CRS of a diesel-dual-fuel engine. The resulting controller is robust to model uncertainties and external disturbances (Chatlatanagulchai, 2010). A quantitative measure of the achieved robustness is also provided and confirmed in simulation via experiments but without considering the effects of cylinder interactions. It is proposed integrator-augmented sliding-mode control on top of existing PIDs controllers with gain scheduling and feed-forward term. This proposed control system was implemented with four cylindrical diesel engines on an engine dynamometer and in a pick-up truck. Using real data, the results compared favorably with the best tuned gain scheduling regular PID controller. Based on that, the possibilities of applying machine learning and digital twin concepts into engine development stages are evidenced. Yet, despite the promising prediction capabilities of the presented works, these models do not support the calibration process and do not anticipate possible faults.

P. Garg et al. provides a comprehensive review of the latest advancements in Machine Learning (ML) techniques used for developing engine control systems, with particular emphasis on the often-lengthy calibration process (Garg, 2021). The review indicates that most research predominantly focuses on regression modeling to capture complex processes, minimize the number of model parameters, and create models suitable for real-time implementation in Electronic Control Units (ECUs). Promising avenues for future research in ML-driven engine control include the use of reinforcement learning to optimize engine performance in real-time and the application of unsupervised learning techniques for monitoring data quality. The author is able to prove that ML techniques can be combined with ECU logic, but unfortunately cannot provide real data to be assessed with the possible implementation.

He, B. et al proposes in his paper a structure and operational mechanism of a digital twin model for tuning PID controllers. By leveraging the capabilities of virtual-real mapping and data fusion provided by the digital twin model, along with online identification of the controlled object's model, the challenges associated with real-time feedback of the controller's actual control effects and the discrepancies in the virtual model due to changing working conditions are effectively addressed. This approach enables the closed-loop self-tuning of the PID controller. Additionally, an intelligent optimization algorithm is proposed to enhance the efficiency and accuracy of the parameter tuning process for the PID controller (He, 2022). The paper also outlines the modeling methodology of the digital twin model across three dimensions: physical prototyping, twin service systems, and virtual prototyping. Finally, the practicality of the proposed method is validated through an example focused on tuning the controller for the stability of gear transmission, without being incorporated in a vehicle ECU.

3.1. Discussion

In this chapter we present the most relevant works in the literature for this research. A summary of the works was structured in Table 1. The works were organized considering the technique used to improve diesel engine calibration process and approaches to their solution, consolidated in the previous chapters. All these selected related works are proposing a solution using digital twin techniques and how it can influence today’s industry.

Table 1 - Related works

	Zheng, 2019	Bhatti, 2021	A. S. Airamadan, 2022	M. Hinrichs, 2021	A. di Gaeta, 2012.	J. Zheng, et al., 2020.	He, 2022	Parson Garg, 2021	W. Chatlatanagulchai, 2010.	**Proposed solution**
Digital Twin framework	X	X						X		**X**
Machine learning application	X	X	X			X	X	X		**X**
Contains a proposed DT/ML solution						X		X		**X**
Application scenarios with examples	X	X	X	X	X	X	X		X	**X**
Real data			X						X	**X**
Proposed solution can influence existing PID controllers				X	X		X		X	**X**
Failure prediction				X		X		X		**X**

The use of digital twin models with machine learning is not new in the automotive industry. As we can also see, there are proposals in the literature for the use of such technologies to assist in the engine calibration process. Unfortunately, until now, access to real and non-simulated data greatly prevents the possible application and implementation of such models in real conditions. The proposed solution in this work contemplates all the possibilities almost reaching a workable solution for the industry, getting this technology closer to being implemented in such cases.

In this way, we summarize the work related to this project, thus concluding the chapter on related work. Continuing the work, Chapter 4 presents the methodological procedures of this work.

Chapter 4

A machine learning-based digital twin model

This section further describes new DT dataset and the proposed machine learning-based digital twin model for the Common Rail System (CRS) of a diesel-engine vehicle. This proposed model, following the DT application framework, is applied in the *operational* stage of the product life cycle.

4.1. Digital-twin Diesel Engine Dataset (DTDED)

This work presents a new DT dataset namely Digital- Twin Diesel Engine Dataset (DTDED). One of the first of its kind, the dataset depicts the data collected from a real diesel-based engine before the release for production. More specifically, DTDED was built through the data collected in one real commercial vehicle with a 2.0 liters diesel engine 4×4 model, with 4 injectors, still in the validation phase. The data collection took place from the controller area network (CAN) network (controller area network) responsible for managing all the information transmitted in the vehicle, from commands sent from the ECU to the reading of data from sensors spread across the other vehicle systems.

In practice, DTDED showcases a miscalibration in a diesel-based engine. The miscalibration occurs due to bad operation by the Common Rail System (CRS), which can result in a Normal, or Overshoot/Undershoot failure situation. Undershoot situations are characterized when the fuel injection system's real pressure (P_{real}) is at least 5% less than the fuel injection system's desired pressure P_{set}. Overshoot situations are characterized when the P_{real} is at least 5% higher than the engine P_{set}. The fuel injection system P_{set} is computed and calculated in the ECU following a well stablished mathematical formula in the automotive industry (Damyot & Chatlatanagulchai, 2013), according to Equation 1.

$$P_{real} = P_{set} + (K_P + K_i + Kd) \quad (1)$$

$$K_P = a * K_{pcrit} \quad (2)$$

$$K_i = b * K_P * 2\pi * f_{crit} \quad (3)$$

$$K_d = K_{pcrit} * (c - a) \quad (4)$$

Where:

K_P	Proportial gain
a	Factor in avoiding K_{pcrit}
K_{pcrit}	Proportional gain value that excites the system at a resonant frequency
K_i	Integral gain
b	Factor in maintaining correlation ship between gains
f_{crit}	Frequency equal or close to natural frequency
K_d	Derivative gain
c	Factor in maintaining correlation-ship between gains
P_{real}	Real fuel injection system pressure (hPa)
P_{set}	Desired fuel injection system pressure (hPa)

The factors (a, b, c) used to calculate P_{set} may vary from system to system according to its configuration and intended usage. (Robert Bosch GmbH, 2006)

This DTDED brings valuable information of not only the CRS operation, but from the entire vehicle and its mode of operation. Every system and boundary conditions of the vehicle can impact the behavior of the CRS, such as the vehicle speed, which correlates not only to the required torque of the engine, but also the amount of inlet air, and therefore amount of oxygen the engine is receiving. A summary with examples of the features can be found in Table 2.

Table 2 - Example of features found in the proposed DTDED

Feature	**Sampling period**	**Description**	**Unit**
AFS_dm	10ms	Sensed fresh air mass flow	kg/s
APP_r	10ms	Standardized accelerator pedal position	%
BattU_u	100ms	Battery voltage	mV
CoEng_st	10ms	Engine operation state (shutoff / running)	0 / 1
EnvP_p	1s	Environmental pressure (measured or modelled)	Pa
Epm_nEng10ms	10ms	Engine speed	RPM
InjCtl_qSetUnBal	10ms	Current fuel injection quantity	mg/injector
Rail_pSetPoint	10ms	Rail pressure setpoint	hPa
VehV_v	100ms	Vehicle speed	Km/h

4.2. Machine learning-based digital twin model

This work proposes a new machine learning-based digital twin model to assist operators with engine calibration and predicting the real pressure P_{set} in the fuel injection system to help the common rail injection system (CRS) to act before a failure occurs. The overall proposal is shown in Figure 5.

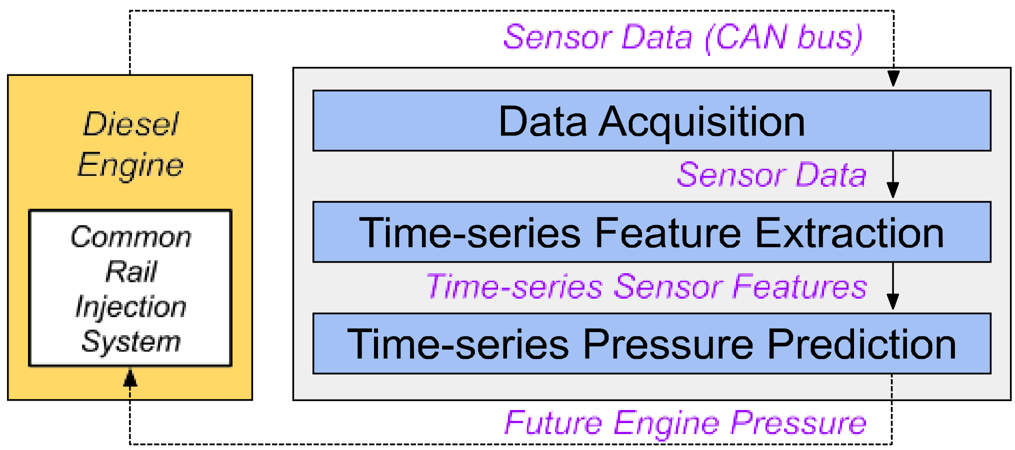

Figure 5 - A digital twin model based on machine learning to predict the pressure in the fuel injection system, helping in the common rail injection system (CRS).

The proposal considers a diesel engine with a CRS module that manages the fuel injection system pressure. The CRS module uses a digital engine twin that continuously evaluates the engine sensors' data to predict pressure level in the fuel injection system.

The main insight of such an approach is that our proposal, based on a digital twin implementation, can predict pressure failure states in the fuel injection system (Overshoot and Undershoot) to assist the CRS module in proactively taking countermeasures. As a result, based on the digital engine twin, the proposal enables the CRS module to avoid situations wherein the fuel injection system pressure may affect the engine's reliability.

The proposal assumes that a diesel engine should not exceed or fall below a preset pressure in the fuel injection system. The engine manufacturer defines the pressure set-point to ensure the reliability of the engine over time. The CRS module must be calibrated to meet the manufacturer's standards, ensuring that the intended pressure levels in the fuel injection system are met. However, the calibration task requires a lot of engineers' time to achieve such a goal. As a result, CRS modules often go to the production line without proper calibration, generating under- and over-pressures in the fuel injection system over time.

4.2.1. Data mining for a digital twin model

This proposal aims to integrate a digital twin model to forecast undesired pressure levels in the fuel injection system. The digital twin aims to replicate the diesel engine behavior as time passes, providing the CRS module with an indicator of undesired pressure levels. The operation of our proposed scheme starts with the data collection by a Data Acquisition module (Figure 5). The module continuously collects engine sensor data through a CAN (Controller Area Network) bus. The collected sensors data values are used as a representation of the diesel engine's current state. The data is used as input by a Time-series Feature Extraction module, which aims to compound a feature vector that depicts the historical behavior of the diesel engine. The module builds a feature vector through a sliding window of events rationale. As a result, the built feature vector depicts the diesel engine state in a given time window, thus, representing the engine behavior as time passes. The built time-series sensor values are used as input by a Time-series Pressure Prediction module. The module, in turn, acts as the digital engine twin to forecast pressure levels in the fuel injection system. To achieve such a goal, it applies a machine learning model, which predicts, within milliseconds in advance, the pressure levels in the fuel injection system. Finally, the CRS module can use the prediction outcome to take counteractions before the pressure level in the fuel injection system reaches undesired states.

Consequently, the proposed machine learning model for building a diesel engine digital twin enables operators to improve the operation of CRS modules. Nevertheless, CRS calibration can be facilitated during the engine development process since calibration inadequacy can be identified and fixed during CRS operations

Data collected in the automotive context follows the Association for the Standardization of Automation and Measurement Systems (ASAM) definitions. Therefore, the data is made available in the Measurement Data Format version 4 (MF4) format, capable of supporting the recording of a high volume of attributes with a high acquisition rate. The data consists of 1,347,340 instances with 208 attributes each. Sample rates range from 1 second, 100 milliseconds, and 10 milliseconds. The collection of DTDED took place in a total period of 10 minutes of vehicle operation.

In the DTDED, feature selection was performed using the information gain approach, measuring the amount of information that one variable (e.g., vehicle speed) provides about another variable (e.g., P_{real}). Specifically, the *mutual_info_regression* function from the scikit-learn library was employed to quantify the mutual dependence between each feature and the target variable. Mutual information measures the amount of information shared between two variables, capturing both linear and non-linear dependencies.

For this analysis, features with a information gain above 0.8 were selected, ensuring that only the most significant features were retained for model training. This threshold was determined empirically to balance the trade-off between model complexity and predictive accuracy, focusing on features that contribute meaningfully to the regression task. This resulted in 70 selected features, as shown in Table 3 – the whole feature list available in the dataset can be found in Chapter 7 - Appendixes. The dataset and the features used in this research are proprietary and owned by engine manufacture. Due to confidentiality agreements and intellectual property restrictions, the dataset and descriptions of what each feature represents cannot be shared publicly.

Table 3 - Feature groups and descriptions

Feature Groups	Description	Number of features
Sensors and actuators of high pressure pump	Data collected from sensor and signals sent from ECU to control the high pressure pump	28
Sensors and actuators of CRS	Data collected from sensor and signals sent from ECU to control the Common Rail System	22
Vehicle drivability	Data collected from sensors regarding driving behavior of the running vehicle	9
Environmental conditions	Data from external sensors in the vehicle. (e.g. external temperature, etc)	4
Injection strategy	Data collected from sensor and signals sent from ECU to control injectors and their strategy. (e.g. moment and duration of fuel injection)	7

4.3. Engine digital twin

A proposal prototype was implemented considering the previously described DTDED dataset. A traditional machine learning process is considered for building our proposed digital twin (Figure 5). The DTDED dataset was split into training, validation, and testing datasets, each composed of 40%, 30%, and 30%, respectively, of the original dataset. This strategy was made possible because it is known that the failures are evenly spread in the dataset. Each dataset depicts the diesel engine operation in a given time window. The dataset sensor values were normalized through a min-max normalization procedure.

Before building our DT model, a linear interpolation procedure was used, given that the dataset was made of 208 feature values collected from a variety of sensors. Each with a sample rate ranging from 1 second (e.g., fuel temperature), to 10 milliseconds (e.g., current amount of injected fuel). The resulted dataset depicts the engine values in a 10-millisecond interval.

Based on the insights provided by the features presented in Table 3 , the DT Time-series Pressure Prediction module was implemented through a linear regression model. The model receives as input a feature vector built by the Time-series Feature Extraction module (Figure 5). The feature vector was built by concatenating the previously selected 70 features considering a 3-sample window.

The model and the previously described data preprocessing were implemented through scikit-learn API v.1.1.1, and pandas API 1.4.2. The model was evaluated through the Root Mean Square Error (RMSE), Adjusted R Square (R2), and Mean Absolute Error (MAE), as usually made in related works. Each of these metrics provides different insights into the model's accuracy and effectiveness.

By using RMSE, R2, and MAE together, a complete evaluation of the model's performance is completed. RMSE highlights larger errors, R2 shows the proportion of variation explained by the model, and MAE provides a direct measure of average error. This combination confirms that the model's accuracy is correctly measured across different aspects, matching both error magnitude and overall fit to the data. (Cabuk, 2023)

Chapter 5

Analysis and evaluation

Current DT datasets used in the literature often do not depict the complexities of the DT domain. This is because as a DT aim at building a digitalized copy of a physical object, the representation of a realistic behavior of the to-be-digitalized object requires the data collection to be made for a significant time, ensuring that even failures can be adequately collected. In contrast, in general, current datasets in the literature either generate the data in a simulated environment or monitor the to-be-digitalized object in unrealistic settings, thus, without data related to failure states.

In this section, we investigate the data distribution of DTDED dataset. More specifically, we first evaluate how the amount of fuel injection from the engine relates to the pressure of the fuel injection system (P_{set}, Eq. 4), as shown in Figure 6. In practice, the engine fuel injection amount is highly correlated with the fuel injection system pressure, reaching a correlation value of 0.94.

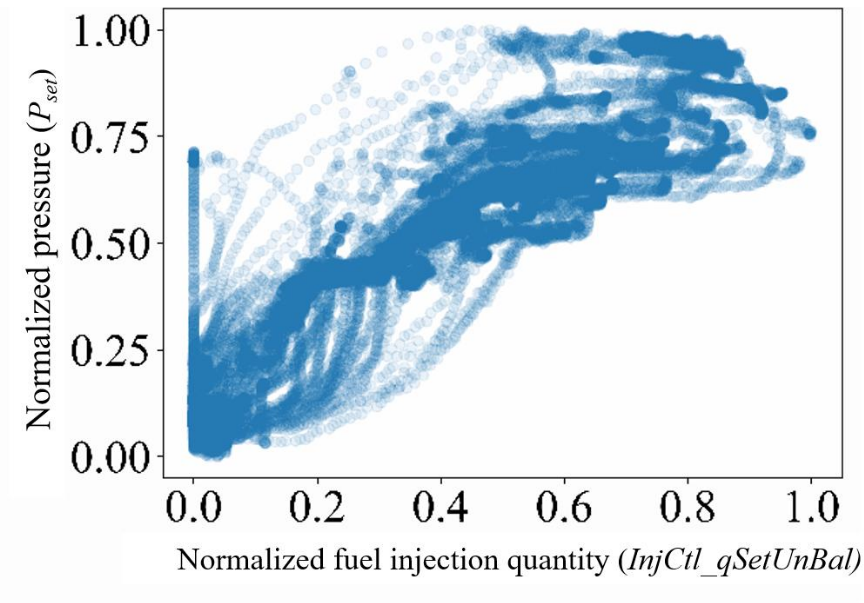

Figure 6 – Data Distribution of DTDED – Normalized fuel injection quantity vs P_{set}

However, it is possible to note a significant dispersion in the collected values caused by the collection of real values from a real diesel-based engine. We further investigate how the failures in DTDED dataset occur. Figure 7 shows a data fraction sample collected from our dataset, showing the occurrence of Undershoot and Overshoot failures as time passes.

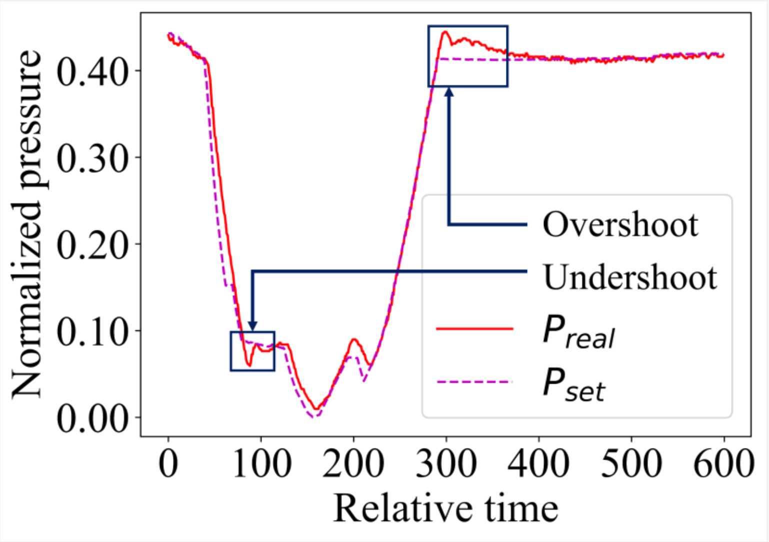

Figure 7 – Data Distribution of DTDED – Undershoot and Overshoot failures

It is possible to note that such a failure continues to occur for a period in time, given the time needed for the P_{real} (real pressure data collected from the vehicle's sensor) to reach the P_{set} (desired fuel injection system pressure being requested by the ECU based on PID calculations). Figure 8 shows the data distribution of our DTDED dataset according to the considered failure states. Considering both Overshoot and Undershoot, failure states account for only 8.14% of the total number of samples.

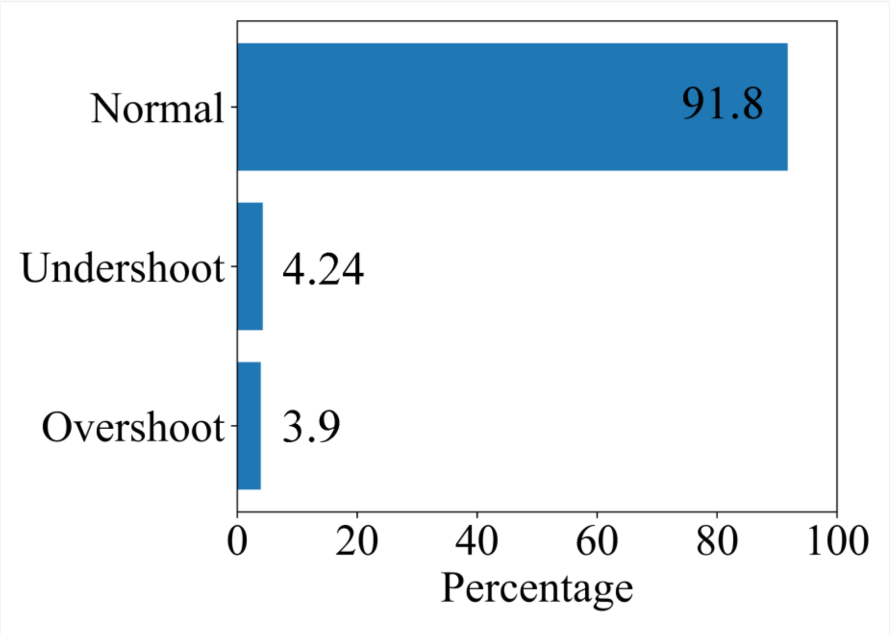

Figure 8 - Data distribution of the DTDED states in percentage

Such data distribution disparity is expected from real diesel engines since failures must occur rarely to ensure proper engine operation. Unfortunately, the rarity of failures affects the construction of realistic DTs, since, in general, prior works relies on collecting data over a small-time window, or even making use of simulated environments. As a result, the DTDED dataset allows operators to construct realistic DTs, given that the engine faults are properly represented in the original dataset. Manufactures do not make it public how much percentage of failures will remain in the engines after the validation phase.

The evaluation aims at answering the following research questions (RQ): (RQ1) How does our proposed DT model work for predicting pressure levels in the fuel injection system? (RQ2) What is the prediction performance of our model for a longer prediction time for the pressure levels?

5.1. A digital twin model

The first experiment aims at answering RQ1 and evaluates the prediction performance of our proposed DT model for pressure levels in the fuel injection system.

We consider a DT model that aims to predict the pressure level in the fuel injection system for the engine CRS 0.1 seconds before it occurs. This prediction time was defined based on calibration expert experience and the physical delay involving physical components like injectors, pumps, and sensors. These components have inherent delays in responding to control inputs due to mechanical and hydraulic limitations, such as the time taken for fuel pressure to change. On top of that, the response time of sensors, which provide feedback to the PID controller, may have latency in detecting changes in pressure or flow rates, causing delays in the controller's ability to react. On top of that, other prediction times will also be evaluated.

The proposal error rates for predicting pressure level 0.1 seconds ahead reached an RMSE of only 0.057, thus, enabling the application of the proposed DT model to assist the CRS module. Figure 9, Figure 10 and Figure 11 show the performance of our proposal in a variety of DTDED dataset settings.

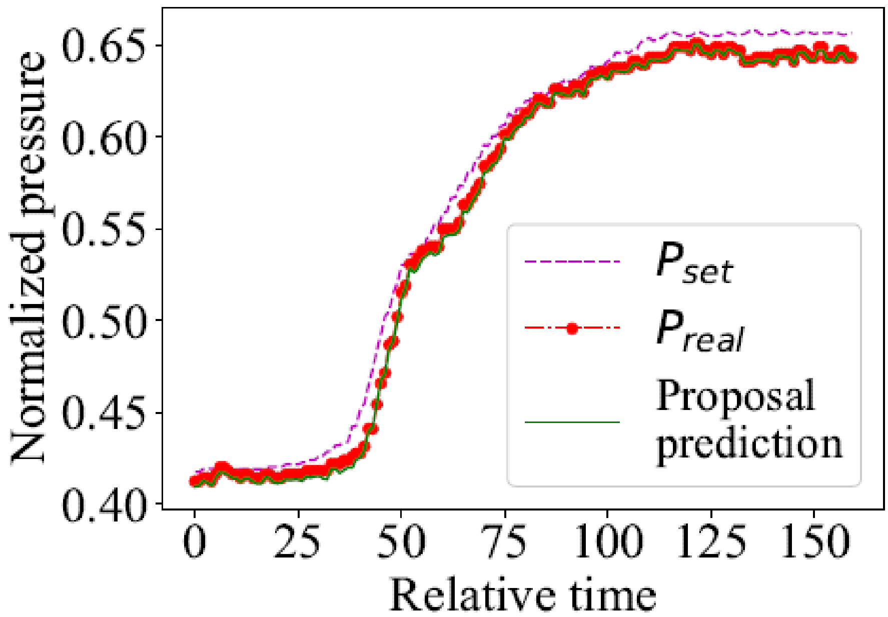

Figure 9 - Proposal performance under different DTDED settings - Normal

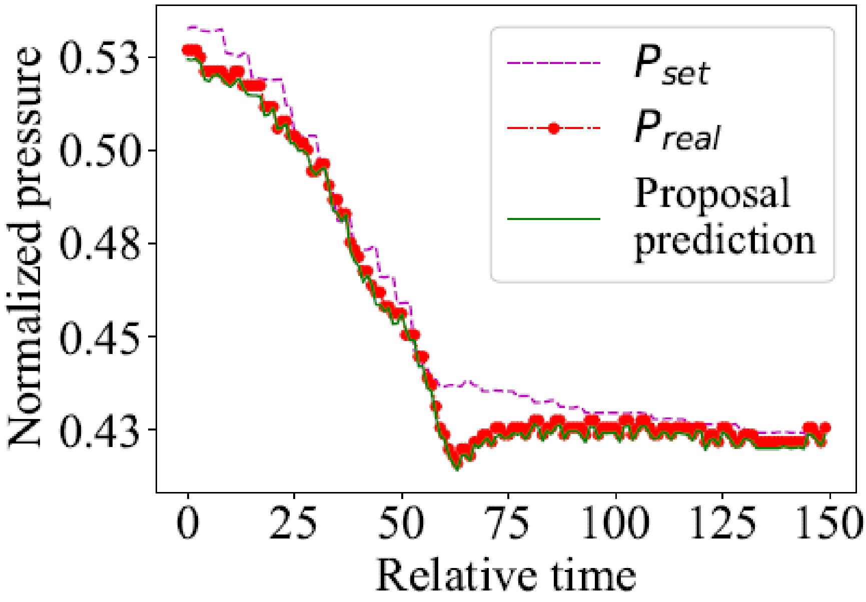

Figure 10 - Proposal performance under different DTDED settings - Undershoot

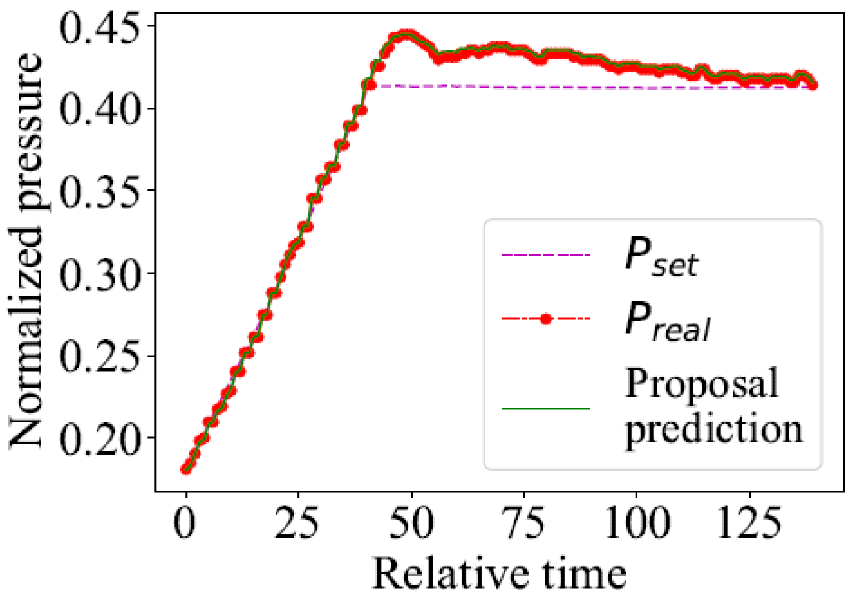

Figure 11 - Proposal performance under different DTDED settings - Overshoot

Our proposed scheme was able to provide similar prediction performance when utilized under Overshoot and above Undershoot to the Normal situation. For instance, in an Overshoot setting, our proposed scheme was able to properly detect the future undesired pressure level in the fuel injection system, with a normalized pressure error of only 0.005 (Figure 11, at ≈ 50 data points). Similarly, our scheme can detect an Undershoot setting with a pressure error of only 0.005 (Figure 10, at ≈ 60 data points).

5.2. Prediction of Pressure Levels in the Fuel Injection System

To answer RQ2, we further investigate how the fuel injection system pressure's prediction time impacts our proposed scheme's prediction performance. Specifically, we vary our model's pressure level prediction time from 0.1 to 0.5 seconds for the fuel injection system. This is because the prediction time for pressure, in the fuel injection system, must be defined according to the expert's needs and may vary according to the used engine configuration. Table 4 shows the prediction performance of our model according to the future prediction time. It is possible to note that the prediction time directly relates to our proposal measured error rates. For instance, the RMSE is increased by 0.006 (+10.6%) when the future prediction time increases from 0.1 to a 0.5 second setting. As a result, the proposed scheme can be by the CRS module to assist in the pressure management for the fuel injection system even if a higher future pressure time is needed. also shows the impact on the prediction performance based on the prediction time in seconds.

Table 4 - Regression performance for predicting P_{real}.

Prediction Time (s)	Root Mean Square Error (RMSE)	R Square (R2)	Mean Absolute Error (MAE)
0.1	0.057	0.949	0.037
0.2	0.058	0.948	0.038
0.3	0.059	0.947	0.039
0.4	0.062	0.941	0.040
0.5	0.063	0.940	0.041

5.3. Prediction of Pressure Levels in PID system

As already mentioned, the common rail system calibration is based on a PID controller embedded in the ECU. The controller's goal to fine-tune the injection system's real pressure can be improved by the usage of the proposed digital twin model. The DT model provides promising results even for half a second in advance, as shown in Figure 12. The DT model can provide input to the PID controller even before the overshoots and undershoots happen, reducing drastically the number of failures in the injection system.

Furthermore, the merging of both models can create an auto tuning PID controller, using the digital twin framework, where no calibration is needed anymore, removing the dependency on experts' knowledge, and reducing drastically the time in engines development and calibration. The merging of both models is a scope not covered in the document but recommended as future works.

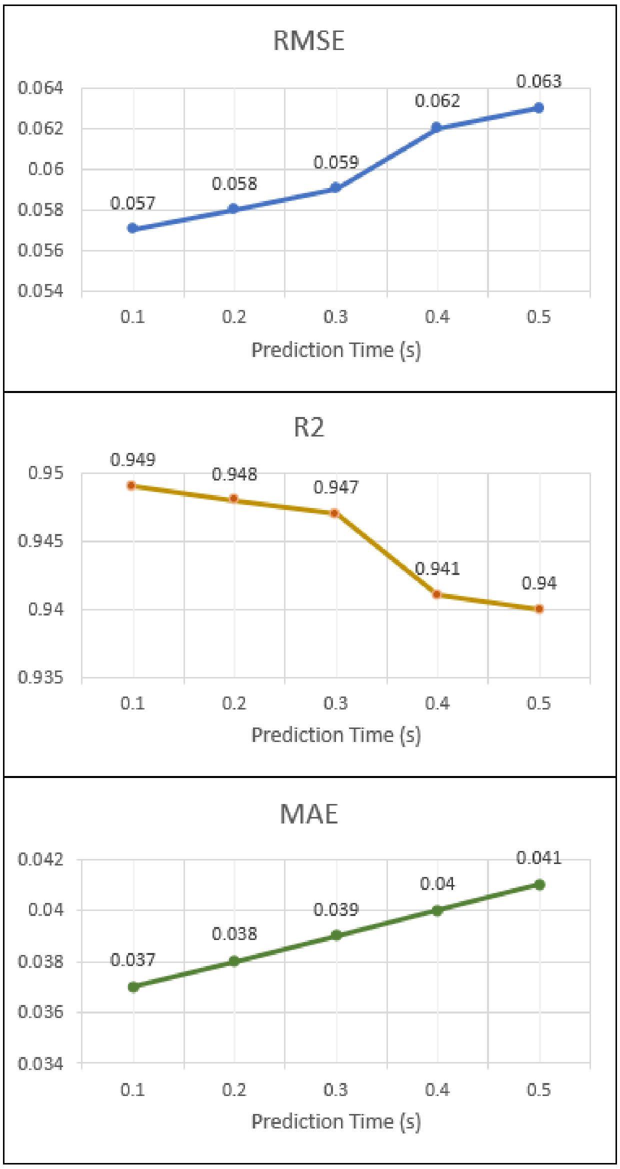

Figure 12 - Regression performance for predicting P_{real} based on prediction time in seconds

5.4. Discussions

This work was able to show the prediction performance of our proposed Digital Twin (DT) model for forecasting pressure levels in a fuel injection system. The DT model aims to predict pressure levels 0.1 seconds before they occur but can also be used for predicting the level 0.5 seconds before. Our results demonstrate that the proposed model achieves a RMSE of just 0.057, showing its efficacy in assisting the CRS controller. Additionally, the model performs consistently across different scenarios, including overshoot and undershoot situations, with minimal normalized pressure errors (0.005). This performance highlights the model's robustness and its potential to mitigate future undesired pressure levels, thereby enhancing the system's reliability.

To further assess the model's effectiveness, we explored how varying the prediction time affects performance. The model remains beneficial for CRS module integration, even under longer prediction intervals. Moreover, the potential for merging this DT model with the PID controller embedded in the ECU presents an exciting future direction.

The method employed in this work proved to be successful, as it achieved the expected results and validated the effectiveness of the DT approach for this specific use case. By accurately predicting the pressure levels in the fuel injection system, the DT model demonstrated its capability to enhance system performance and reliability. The results emphasize the value of Digital Twin technology in providing real-time insights and predictive capabilities that can prevent system failures and optimize operations. This success highlights the potential of DT models to revolutionize fuel injection systems and similar complex engineering applications, showcasing their practical utility and transformative impact.

Chapter 6

Conclusion

Digital twin is the critical technology to fully fusion physical and virtual models. A technology that has been pursued for many years, but based on its complexity, it is still not common to see fully implemented frameworks in industries and final products. This work has proposed a digital twin model based on machine learning techniques to assist the vehicle engine development task. PIDs controllers are the barebone of the calibration processes of the main vehicle functions, and it has been proven reliable once properly used, but a time-consuming task that cannot be automatized by itself.

The proposed scheme predicts pressure levels in the fuel injection system of a real diesel-based vehicle engine. It was shown that many data mining and machine learning techniques needs to be used to proof the digital twin value and impact, nevertheless, as a result, it can be used to assist the Common Rail System (CRS) module in preventing unwanted pressure levels in the fuel injection system, avoiding the engine from premature wear.

Despite these challenges, the integration of a PID controller within the digital twin framework offers unprecedented opportunities for advancing diesel engine calibration and control. The usage of data mining and machine learning techniques are crucial for DT models, especially with time-series data. Furthermore, by incorporating digital twin results into the control loop, future researchers can achieve continuous improvement, proactive maintenance, and performance optimization, thereby realizing the full potential of digital twin technology to revolutionize diesel engine development and operation in the automotive domain.

As future works, we plan to extend the proposed dataset, not only to work with more than 10 minutes of data collection, but also to include additional vehicles and incorporate the proposal in an actual engine, using the Digital Twin (DT) model as input for the PID controller. This proposal can be extended in not only a reduction of overshoots and undershoots, but also a “non-fixed” calibration, adapting to external environments and different boundaries.

Due to the collaborative nature of this work with engine manufacturer, data and information related to the investigation and development cannot be disclosed. Part of the content is sensitive and classified as confidential, as it involves proprietary knowledge and protected intellectual property. These restrictions are in place to honor the confidentiality agreements established with the company.

Chapter 7

Appendixes

Appendix A – DTDED Features

Feature groups	Features	Frequency	Description
Environmental conditions	*Altitude*	1s	-
Environmental conditions	*EnvP_p*	1s	Env. pressure, measured or modelled
Environmental conditions	*EnvT_t*	1s	Environment temperature signal (with replacement reaction in case of sensor error)
Environmental conditions	*FuelT_t*	1s	Temperature of fuel
FIE	*CEngDsT_t*	1s	Coolant temperature at engine output
FIE	*ETCtl_qPoI1*	10ms	Additive correction of retarded post injection quantity
FIE	*InjCrv_phiMI1Des*	10ms	Desired reference angle for the start of MI1
FIE	*InjCrv_phiMI1Set*	10ms	Desired SOD of MI1
FIE	*InjCrv_qMI1Des*	10ms	Desired MI1 injection quantity
FIE	*InjCrv_qPiI1Des_mp*	10ms	Desired quantity for PiI1
FIE	*InjCrv_qPiI2Des_mp*	10ms	Desired quantity for PiI2
FIE	*InjCrv_qPiI3Des_mp*	10ms	Desired quantity for PiI3
FIE	*InjCrv_qPoI1Des_mp*	10ms	Post injection 1 setpoint quantity
FIE	*InjCrv_qPoI2Des_mp*	10ms	Post injection 2 setpoint quantity
FIE	*InjCrv_qPoI3Des_mp*	10ms	Post injection 3 setpoint quantity

FIE	*InjCtl_qCurr*	10ms	Torque generating engine fuel injection quantity
FIE	*InjCtl_qSetUnBal*	10ms	Current injection quantity
FIE	*MeUn_dvolAdaptCrvCor*	100ms	Adaptive correction of metering unit set point value
FIE	*MeUn_dvolSet*	100ms	Setpoint value (volume flow) of rail pressure governing
FIE	*MeUn_dvolSetAdapt*	100ms	Correction value of the adaption of the metering unit
FIE	*Rail_pSetPoint*	10ms	Rail pressure setpoint
FIE	*RailP_pFlt*	10ms	Maximum rail pressure of the last 10ms
FIE	*RailP_uRaw*	100ms	Raw value of rail pressure
Inlet AIR SYSTEM	*AFS_dm*	10ms	Sensed fresh air mass flow
Inlet AIR SYSTEM	*AFS_facAdjVal_[0]*	1s	Final Correction factor for the air mass for drift stored in EEPROM
Inlet AIR SYSTEM	*AFS_facAdjVal_[1]*	1s	Final Correction factor for the air mass for drift stored in EEPROM
Inlet AIR SYSTEM	*AFS_mAirPerCyl*	10ms	Air mass per cylinder
Inlet AIR SYSTEM	*AFS_mAirPerCylFlt*	10ms	Filtered air mass per cylinder
Inlet AIR SYSTEM	*AirCtl_mDesBasCor*	10ms	Calibration for desired air mass Base Correction
Inlet AIR SYSTEM	*AirCtl_mDesVal*	10ms	Desired air mass value for low temperature combustion
Inlet AIR SYSTEM	*AirCtl_rTVAClsdLoop_mp*	100ms	Ratio throttle valve Closed Loop
Inlet AIR SYSTEM	*AirCtl_stMon*	10ms	Status: shutdown case of the governor
Inlet AIR SYSTEM	*ASMod_dmIndAirRef*	10ms	Reference gas mass flow into the engine
Inlet AIR SYSTEM	*PCR_pActVal*	100ms	Boost pressure actual value
Inlet AIR SYSTEM	*PCR_pDesVal*	100ms	Limited boost pressure setpoint
Inlet AIR SYSTEM	*ThrVlv_rAct*	100ms	Actuator position
Inlet AIR SYSTEM	*ThrVlv_rDesVal*	100ms	-
Inlet AIR SYSTEM	*TrbCh_rAct*	100ms	Actuator position
Inlet AIR SYSTEM	*TrbCh_rDesVal*	100ms	Desired position

Outlet AIR SYSTEM	*ASMod_dmEGFld_[0]*	100ms	Air mass flow for exhaust upstream pressure calculation
Outlet AIR SYSTEM	*DewDet_wLSU_[0]*	1s	Integrated heat quantity at sensor position.
Outlet AIR SYSTEM	*DewDet_wLSU_[1]*	1s	Integrated heat quantity at sensor position.
Outlet AIR SYSTEM	*LSU_rO2Act_[0]*	100ms	Actual oxygen value without filter and without freeze, only offset compensated
Outlet AIR SYSTEM	*LSU_rO2Adap_[0]*	100ms	Adapted O2 ratio after over run correction
Outlet AIR SYSTEM	*LSU_tRi_[0]*	1s	Temperature of lambda sensor
Outlet AIR SYSTEM	*SmkLim_qLimSmk*	100ms	ramped smoke limitation quantity
Vehicle states	*BattU_u*	100ms	Battery voltage
Vehicle states	*T15_st*	10ms	Terminal 15 status after debouncing
Vehicle states	*T50_st*	1s	Status of T150 signal
Vehicle states	*ActMod_trqClth*	10ms	actual engine torque - clutch torque
Vehicle states	*APP_r*	10ms	Standardized accelerator pedal position
Vehicle states	*APP_uRaw*	10ms	Acceleration pedal position raw value
Vehicle states	*Epm_nEng*	100ms	Average engine speed of one cylinder segment
Vehicle states	*Epm_nEng10ms*	10ms	Engine speed calculated in 10ms
Vehicle states	*Tra_numGear*	100ms	Current gear information
Vehicle states	*Brk_st*	100ms	Brake switch state
Vehicle states	*Brk_stMn*	100ms	State main brake switch
Vehicle states	*Brk_stRed*	100ms	State redundant brake switch
Vehicle states	*CoEng_st*	10ms	Value of CoEng_st if engine is running
Vehicle states	*CoEOM_numStageAct*	10ms	Number of the active stage of the operation mode
Vehicle states	*CoEOM_stOpModeAct*	10ms	Active operation mode
Vehicle states	*CoETS_stCurrLimActive*	10ms	Status of active minimum of limitation torques

Vehicle states	*DnoxCtl_flgEngStopEna*	10ms	Engine stop enable by DnoxCtl
Vehicle states	*DnoxCtl_tiEngOff*	1s	DnoxCtl internal engine-off-time
Vehicle states	*EngDa_tiEngOn*	1s	Engine on time
Vehicle states	*Epm_stOpMode*	10ms	State of EPM operation mode
Vehicle states	*GlbDa_lTotDst*	1s	Total distance since first start
Vehicle states	*SyC_stMn*	100ms	Current system/ECU state
Vehicle states	*VehV_v*	100ms	Vehicle speed
Vehicle states	*DewDet_stHtgLSUHeat_mp*	100ms	Status whether heat quantity exceeds heat threshold value after a certain delay

References

A. S. Airamadan, Z. A. I. B. M. a. J. B., 2022. Machine learning model for spark-assisted gasoline compression ignition engine. *SAE International,* Volume SAE Technical Paper Series.

Abreu, A. O. S. E. K. V. a. V. V. C., 2020. Identity and access management for IoT in smart grid. *Advanced Inf. Net. and App. Springer International Publishing,* p. pp. 1215–1226.

Anderl, S. H. a. R., 2018. *Digital twin – proof of concept.* vol. 15, pp. 64–66, Industry 4.0 and Smart Manufacturing. ed. s.l.:Manufacturing Letters.

Anon., 2021. *Digital Twin: Apply Advanced Analytics and Machine Learning to Reduce Operational Costs and Risks.* [Online] Available at: https://www.ge.com/digital/applications/digital-twin?utm_medium=Paid-Search&utm_source=Google&utm_campaign=HORZ-DigitalTwin-MoF-EU-Search&utm_content=%2Bdigital%20%2Btwin
[Accessed 20 January 2024].

Bhatti, G., 2021. *Towards the future of smart electric vehicles: Digital twin technology.* vol. 141, p. 110801 ed. s.l.:Renewable and Sustainable Energy Reviews.

Bulle, A. O. S. E. K. V. a. R. R. d. S., 2020. A host- based intrusion detection model based on OS diversity for SCADA.

Cabuk, S. K. C. M. G. G. e. a., 2023. Chasing the objective upper eyelid symmetry formula; R2, RMSE, POC, MAE and MSE. *PREPRINT*, 29 September.

Carpenter, J. et al., 2018. Machine Learning at a Gold-Silver Mine: A case study from the Ban Houayxai Gold-Silver Operation. *In Proceedingsof the Complex Orebodies Conference 2018,* November, pp. 82-85.

Chatlatanagulchai, W., 2010. Gain-scheduling integrator-augmented sliding-mode control of common-rail pressure in diesel-dual-fuel engine. *SAE Technical Paper Series. SAE International*.

Chen, T. L. J. &. S. B., 2024. An Exploration of the Development of Computerized Data Mining Techniques and Their Application.. *International Journal of Computer Science and Information Technology,* Volume 3, pp. 206-212.

Damyot & Chatlatanagulchai, 2013. Common-Rail Pressure Control in a Diesel-Dual-Fuel Truck Using Sliding Mode. *The 4th TSME International Conference on Mechanical Engineering*, Issue DRC-1021.

Deffo, B. B. A., 2024. Robust frequency control based on sliding mode control with LMI in an island microgrid.. *International Journal of Dynamics and Control.*

Gaeta, A. d., 2012. A model-based gain scheduling approach for controlling the common- rail system for GDI engines. *International Journal of Control,* vol. 85(no. 4), p. pp. 419–436.

Garg, E. S. F. W., 2021. Potential of Machine Learning Methods for Robust Performance and Efficient Engine Control Development. *IFAC-PapersOnLine,* 54(10), pp. 189-195.

Glaessgen, E. &. S. D., 2012. *The digital twin paradigm for future NASA and U.S. air force vehicles.* s.l., 53rd Structures, Structural Dynamics and Materials Conference.

Goodwin, T. X. J. C. N. &. C. C. H., 2022. Real-time digital twin-based optimization with predictive simulation learning.. *Journal of Simulation,* Volume 18, pp. 47-64.

Grieves, M., 2016. *Origins of the Digital Twin Concept.* [Online]
Available at: https://www.researchgate.net/publication/307509727_Origins_of_the_Digital_Twin_Concept
[Accessed 2024].

He, B. L. T. a. X. J., 2022. Digital Twin-Driven Controller Tuning Method for Dynamics. *ASME. J. Comput. Inf. Sci. Eng,* 25 March, pp. 21-24.

Hinrichs, M., 2021. *Online fault detection of a heavy duty diesel engine with model-based methods.* [Online]
Available at: https://tuprints.ulb.tu- darmstadt.de/id/eprint/18886
[Accessed 3 Jun 2024].

Horchulhack, E. K. V. a. A. O. S., 2022. *Toward feasible machine learning model updates in network-based intrusion detection.* vol. 202, p. 108618, ed. s.l.:Computer Networks.

Kameshwaran, K. a. K. M., 2014. Survey on clustering techniques in data mining.. *International Journal of Computer Science and Information Technologies,* pp. 2272-2276.

Knecht, W., 2008. *Diesel engine development in view of reduced emission standards.* vol. 33, no. 2, pp. 264–271 ed. s.l.:Energy.

Kotsiantis, S. B. D. K. a. P. E. P., 2006. Data preprocessing for supervised leaning.. *International journal of computer science,* pp. 111-117.

Laubichler, C. K. S. W. M. K. A. W. M. C. C. N. D. L. H.-M. K. T. K., 2024. Impact of Injection Valve Condition on Data-driven Prediction of Key Combustion Parameters Based on an Intelligent Diesel Fuel Injector for Large Engine Applications. *SAE MOBILUS,* Volume 0148-7191, p. 16.

Liu, L. L. R. C. K. O. T. F. L. S. B., 2022. Digital Twin-enabled Collaborative Data Management for Metal Additive Manufacturing Systems. *Journal of Manufacturing Systems,* 62(ISSN 0278-6125), pp. 857-874.

Lu, J. Z. V. M. T. W., 2024. Study on precise fuel injection under multiple injections of high pressure common rail system based on deep learning. *Energy,* 307(132784).

Lygoe, R. J., 2010. *A many-objective optimisation decision-making process applied to automotive diesel engine calibration.* pp. 638–646. ed. s.l.:Lecture Notes in Computer Science.

Meli, Z. W. P. B. S. P., 2024. Neural Network Modeling of Black Box Controls for Internal Combustion Engine Calibration. *SAE MOBILUS,* p. 14.

Nicoleta, I., 2011. Perspectives of inovation in Automotive Industry and Costumer Feedback for the new Technologies. *ModTech International Conference,* pp. 25-27.

Oyekan, J. et al., 2021. Applying a 6 DoF RoboticArm and Digital Twin to Automate Fan-Blade Reconditioning for Aerospace Maintenance,Repair, and Overhaul.. *Sensors,* Volume 4637, p. 20.

Ramos, et al., 2021. *A machine learning model for detection of docker-based APP overbooking on kubernetes.* s.l.:ICC 2021 - IEEE International Conference on Communications.

Robert Bosch GmbH, 2006. *Diesel Engine Management.* s.l.:John Wiley and Sons Ltd.

Sajal, M. R. I. M. M. I. F. R. B. M. F. A. M. H. A. S. H. A. S. K. S. M. M. H., 2024. Digital twin: Data exploration, architecture, implementation and. *Heliyon.*

Sandberg, A., 2024. *Discovering technologies in the automotive industry,* Enschede - Netherlands: BMS: Behavioural, Management and Social Sciences - University of Twente.

Schluse, M., 2018. *Experimentable digital twins—streamlining simulation-based systems engineering for industry 4.0.* vol. 14, no. 4, pp. 1722–1731 ed. s.l.:IEEE Transactions on Industrial Informatics.

Shafto, M. et al., 2010. *DraftModeling, Simulation, Information Technology & Processing Roadmap..* [Online]
Available at: https://www.nasa.gov/pdf/501321main_TA11-MSITP-DRAFT-Nov2010-A1.pdf
[Accessed 05 February 2024].

Shao & Helu, M., 2020. Framework for a digital twin in manufacturing: Scope and requirements. *Manufacturing Letters,* Volume ISSN 2213-8463, pp. 105-107.

Singh, M. et al., 2021. Digital Twin: Origin to Future. *Applied System Innovation,* Volume 4, p. 36.

Singh, M. et al., 2022. Applications of Digital Twin across Industries: A Review.. *Appl. Sci,* Volume 5727, p. 12.

Soltanalizadeh, E. V. H. Y. M. N. M., 2024. Automatic smooth map generation of internal combustion engines via local-global model based calibration technique.. *International Journal of Engine Research.,* Volume 25, pp. 990-1012.

Tao, F. C. J. Q. Q. e. a., 2018. Digital twin-driven product design, manufacturing and service with big data. *Int J Adv Manuf Technol* , Volume 94, pp. 3563-3576.

Viegas, Santin, A. O., Cogo, V. V. & Abreu, V., 2020. A reliable semi-supervised intrusion detection model: One year of network traffic anomalies. *IEEE Int. Conf. on Comm. (ICC),* p. pp. 1–6.

Xie, R. et al., 2021. Digital Twin Technologies for Turbomachinery in a Life Cycle Perspective: A Review. *Sustainability,* Volume 13, p. 2495.

Xi, W., 2018. *A feature extraction and visualiza- tion method for fault detection of marine diesel engines.* vol. 116, pp. 429–437 ed. s.l.:Measurement.

Ye, Y. et al., 2020. Digital Twin for the Structural HealthManagement of Reusable Spacecraft: A case study. *Eng. Fract. Mech.,* Volume 107076, p. 234.

Zheng, J., 2020. Fault diagnosis of injector in high pressure common rail system of diesel engine. *11th International Conference on Prognostics and System Health Management (PHM-2020 Jinan).*

Zheng, Y., 2019. An application framework of digital twin and its case study. *Journal of Ambient Intelligence and Humanized Computing* .