IECON 2025 – 51st Annual Conference of the IEEE Industrial Electronics Society | 979-8-3315-9681-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/IECON58223.2025.11221373

Exploring Automotive Quality Correlations through Explainable Machine Learning What-If Simulation

Alexandre O. Júnior*†, José Luis Calvo-Rolle†‡, Rui Pires§, Paulo Leitao*
* Research Centre in Digitalization and Intelligent Robotics (CeDRI), Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal, Email: {alexandrejunior, pleitao}@ipb.pt
†Universidad de La Laguna, Dep. de Ingeniería Informática y de Sistemas, La Laguna, 38204, Tenerife, Spain
‡Department of Industrial Engineering, University of A Coruña, CTC, CITIC, Rúa Mendizábal, s/n, 15403 Ferrol, A Coruña, Spain, E-mail: jlcalvo@udc.es
§Volkswagen Autoeuropa, 2954-024 Quinta do Anjo, Portugal, Email: rui1.pires@volkswagen.pt

***Abstract*—High-dimensional variability in manufacturing processes presents significant challenges for quality control, demanding predictive strategies capable of capturing complex parameter dependencies. Machine learning (ML) offers robust mechanisms for this purpose, but reliance on black-box models often limits interpretability and hinders producing stakeholders' identification of meaningful correlations for model optimization. This paper introduces an interactive what-if simulation platform designed to explore structural quality correlations in automotive assembly through explainable ML techniques, enhancing transparency and enabling uncertainty quantification. The platform is based on a modular Digital Twin (DT) architecture aligned with the ISO 23247 standard, guiding expert and non-expert users through correlation-driven feature selection, regression modelling and SHapley Additive exPlanations (SHAP) based post-hoc explanations. A case study using real inspection data from a vehicle assembly line demonstrates the tool's capacity to support variable relevance assessment, dimensionality reduction, and model interpretability. Furthermore, an uncertainty-aware SHAP analysis enhances confidence in the model's prediction stability, reinforcing the platform's suitability for quality-driven decision support and integration into future DT ecosystems.**

Keywords: Explainable Machine Learning, Correlation Analysis, Uncertainty-aware, Simulation, Industrial Metaverse, Digital Twin

I. Introduction

Industry 4.0 has been promoting the digital transformation of manufacturing by adopting cutting-edge technologies such as Cyber-Physical Systems, Artificial Intelligence (AI), and Digital Twin (DT), fostering seamless process monitoring and interconnection of devices across the shop floor [1]. DTs allow the creation of realistic virtual replicas of physical assets to enhance control and decision-making based on real-time data, predictive analytics, and what-if simulations. As these technologies mature, the Industrial Metaverse paradigm is emerging. It aims to expand the interconnectivity of processes through DT's networks that span the entire production lifecycle. In addition, it emphasizes human participation in industrial activities in a more immersive, intelligent, and resilient manner — aligning with the Industry 5.0 (I5.0) principles [2].

Optimizing industrial systems by simulating and evaluating diverse what-if scenarios is a core functionality of DTs, which enables predictive analysis and supports the implementation of Zero Defect Manufacturing (ZDM) strategies. Within the Industrial Metaverse, these simulations can be combined with AI techniques to improve model accuracy, generate synthetic data, and accelerate computational processes [3]. This integration also enables extracting meaningful patterns and identifying critical variables from operational data, uncovering causal relationships that influence system behaviour. Furthermore, DTs can incorporate expert insight and user feedback through iterative loops, fostering continuous model refinement and the development of adaptive learning strategies that raise both the reliability and contextual relevance of simulations [4].

Integrating expert knowledge and AI becomes even more valuable in complex industrial scenarios, such as automotive quality control. Vehicle assembly quality depends on maintaining multiple measurement points within defined tolerance limits. However, the data's high dimensionality and non-linearity in these environments often hinder the construction of reliable models that explain correlations among critical variables — especially when relying exclusively on traditional statistical methods. AI and Machine Learning (ML) approaches offer more robust and accurate modelling capabilities. However, their application remains challenging and subject to uncertainty, as their performance strongly depends on the quality, completeness, and relevance of the data. Redundant features can introduce noise and lead to overfitting, while missing key variables may prevent the model from capturing essential behaviours, compromising its predictive reliability [5].

Adopting explainable ML techniques is a promising strategy to quantify uncertainty in industrial systems, improving the transparency and trustworthiness of integrated AI models in DTs. Explainability allows users to understand the impacts and contributions of individual input features of the model's predictions, helping to identify and exclude irrelevant or misleading variables and better interpret existing correlations among them. Furthermore, explainable mechanisms support a two-way learning dynamic in which models are continuously refined through expert knowledge while experts and non-experts gain deeper insights regarding system behaviour, as well as ensuring higher confidence in the models for decision-making. However, translating explainable models into DTs to

manage data uncertainty remains an unsolved challenge [6].

This paper presents an explainable what-if simulation platform for exploring correlations designed to support dimensional quality assessment in the automotive industry. Through a visual and feedback-driven interface, the platform enables users to interactively explore and quantify correlations between measurement points by combining statistical analysis with explainable ML models tailored to different dependency relationships. This approach aims to contribute to bridging the gap between DT data-driven modelling and engineering expertise in critical quality processes. The platform's application to real-world automotive quality data demonstrates its potential to support model transparency, interpretability, uncertainty awareness, and expert-in-the-loop refinement.

The rest of the paper is organized as follows: Section II briefly analyses the ML explainability in DTs, highlighting their relationships and current gaps. Section III presents an architecture aligned with the ISO 23247, designed to identify correlations among quality process parameters by integrating explainable ML and exploratory analysis into a DT-based what-if simulation tool. Section IV describes the automotive measurement quality case study. Section V details the tool implementation, highlighting the correlation analysis workflow, regression modelling, and model explainability. Section VI describes and discusses the simulation tool's preliminary experimental validation. Finally, Section VII rounds up the paper with the main conclusions and points out future work.

II. ML EXPLAINABILITY IN DT

The convergence of the Industrial Metaverse with DT technologies is paving the way for advancing AI-driven simulation methodologies for manufacturing decision-making. Recent approaches highlight the role of human-centric systems, where immersive environments enable experts to interact with real-time data and simulation outcomes [2], [7]. However, establishing trustworthy AI for DTs in a context-aware fashion remains a significant challenge. Typically, AI algorithms applied to classification or regression tasks rely on black-box models that produce estimations without explanations or uncertainty measures, inhibiting confidence in the predictions [8].

Incorporating explainable ML approaches enables a more transparent interpretation of the model's outputs and internal computation, including the weight distribution of input variables coefficients. This explanation process helps users identify model projections' relative strengths and weaknesses, supports intuitive expert-driven tunning, facilitates clear parameter selection and configuration, and contributes to reducing overfitting and sensitivity to input perturbations. Regardless, existing explainable ML techniques still face limitations that can lead to misleading explanations, making their effective integration into DTs even more challenging [9].

ML explainability relies on model simplification, feature importance attribution, post-hoc analysis, and human-in-the-loop integration. Feature selection procedures are paramount for model simplification, as they judiciously select feature subsets, mitigate data noise and multicollinearity, lower training and computing costs, and streamline intuitive data visualization. Still, the semantic relevance of selected features is often overlooked in ML-based systems. Although expert-augmented feature selection remains underexplored in literature, emerging approaches showcase the potential to enhance transparency and justifiability across all phases of the modelling pipeline by aligning dimensionality reduction decisions with domain knowledge and contextual priorities [10].

Feature importance can be effectively addressed using SHapley Additive exPlanations (SHAP) [11], a game-theoretic algorithm based on Shapley values that quantifies the contribution of each input feature across all possible subsets. Unlike traditional importance metrics, SHAP captures complex interactions between features, including both positive and negative influences on model output. Within production environments, this approach can identify the variables that contribute to the occurrence of defects, combining interpretability alongside the predictive strength of black-box ML models and enabling the fine-tuning of control parameters for defect reduction [12].

SHAP's ability to identify individual features' contribution to the model's outputs makes it well-suited for post-hoc explanations, offering interpretable insights that benefit expert and non-expert users. To our knowledge, few studies have explored these post-training explanations in the DT context. However, a notable example is the ENIGMA platform [13], which uses DT-based simulations combined with SHAP to explain AI-driven decisions in attack prediction tasks, enabling security stakeholders to interpret model behaviour and take preventive actions. This integration exemplifies how explainability in DTs can support the human-centric and resilient principles of I5.0.

In complex multistage production scenarios, such as the automotive industry, conditions such as the high-dimensional variability of the vehicles facilitate the downstream propagation of defects. Tree-based regressors such as Random Forest and Extreme Gradient Boosting (XGBoost) are particularly valuable for quality control, given their capacity to handle multivariate, high-dimensional data and capture non-linear relationships, which is essential given the inherent uncertainties and disturbances in the assembly process [14]. However, unlike traditional exploratory analysis methods based on correlation matrices, which are more easily interpretable, these regressors tend to overshadow engineering understanding and hinder the development of predictive approaches. This challenge underlines the significance of employing explainable ML strategies, not only to provide insights to domain experts but also to actively incorporate their experience and knowledge of the production process to improve the model's development and correlation analysis, especially since purely statistical approaches may lead to misleading correlations [15].

This paper delves into the search for correlations in automotive dimensional quality by combining exploratory correlation analysis with regression models, supported by an interactive what-if simulation interface and explainable ML techniques to enhance the investigation of intrinsic dependencies among quality measurement points, leveraging the construction and refinement of predictive models through expert feedback.

III. Architecture for Explainable ML-Enhanced What-If Simulation for Correlation Exploring

Intending to operationalize the proposed approach and offer some design guidelines for integrating explainable ML into I5.0 scenarios, Figure 1 presents a generic architecture for user-oriented correlation exploration in quality control. The architecture was designed in compliance with the ISO 23247 standard, which defines an entity-based reference framework for DTs in manufacturing environments [16].

The **Manufacturing Assets** layer comprises the physical assets to be digitalized, in this case, all the equipment responsible for performing the quality inspections. Within the *Industrial Metaverse Framework*, the **Communication Entity** handles the data acquisition from these shop floor quality inspection assets and stores it in a raw database. A *data preprocessing* module then cleans and filters the collected data into a structured format suitable for downstream modelling. The *user feedback sync* module represents the bidirectional flow of domain knowledge and user decisions between the interface and the DT applications.

The **User Entity** hosts the *simulation interface*, through which the user interacts with the what-if framework to define correlation analysis targets, explore intermediate results, and validate model behaviour through human-in-the-loop interactive knowledge exchange. The interface must support expert and non-expert users by providing visual explanations and intuitive configuration mechanisms.

The core correlation analysis functions are in the **Digital Twin Entity**, where the *processed database* feeds a structured pipeline for exploring dependencies via the *explainable ML what-if simulation tool*. In this environment, the user can simulate and evaluate different input variable subsets for predictive model construction according to the correlation with the target variable. This pipeline comprises the following modules:

- **Exploratory correlation analysis**: In this initial stage, correlations between the selected target variable and the remaining dataset are computed, accounting for both linear and non-linear relationships.
- **User-guided feature selection**: The interface displays the computed correlations (e.g., using heatmaps or scatter plots), allowing users to visualize potential relationships between variables. Drawing on their production process knowledge and guided by adjustable correlation thresholds, users can make a preliminary selection of the input features for model construction. This step also benefits non-expert users by providing visual guidance for contextualizing statistical patterns and facilitating informed decisions even without in-depth process knowledge.
- **Regression modelling**: Based on selected input features, predictive models for the target variable tailored to each correlation type are generated (e.g., Linear Regression for linear relationships, Random Forest for non-linear).
- **Feature selection**: The models undergo dimensionality reduction by feature selection techniques (e.g., forward/floating selection), which optimize predictive performance by eliminating irrelevant or misleading features.
- **Post-hoc explanation**: Explainable ML algorithms are applied to quantify the contributions of each input feature to the model's predictions, providing interpretability and transparency to the users regarding how the model generates its outputs.
- **Final model selection**: Based on the explanations provided, the individual performance scores for each model, and practical relevance, the user can select the most appropriate model to be stored.

As a result of this pipeline, the *selected predictive model* can be deployed in various *decision support tools* for predictive monitoring, continuous quality improvement, and other what-if applications, serving as a validated and trustworthy model within the broader DT ecosystem.

IV. Automotive Measurement Quality Case Study

An automotive company's assembly line's quality inspection process was adopted as a case study to validate the proposed architecture. As previously mentioned, the assembly process consists of multiple sequential stages in which strict dimensional quality tolerance limits must be met to ensure compliance with production standards, especially for structural points classified as critical due to their high impact on the vehicle's overall quality.

Figure 2 illustrates the assembly and inspection operations in the automotive body shop, covering the *body-in-white* phase, which precedes vehicle painting and final assembly. The body shop is organized into three distinct assembly zones, which are managed by multiple automated stations dedicated to the vehicle's framing, doors and tailgate assembly. At the end of each stage, robotic stations equipped with laser triangulation systems perform dimensional inspections of multiple quality points. The framing zone involves the initial welding and assembly of stamped components, where only geometrical features (X, Y, and Z coordinates) and their respective symmetries are available for inspection. These measurements are compared against the vehicle's reference CAD model to detect geometric deviations. In contrast, the subsequent doors and tailgate zones already include assembled parts, enabling the inspection regarding the measurement of gap and flush features, which respectively assess the distance and alignment between adjacent parts.

ML is pivotal for establishing predictive quality strategies throughout the body shop by modelling complex dependencies between early-stage dimensional measurements and downstream defect occurrences, leveraging large volumes of inspection data to anticipate potential quality issues before they manifest, supporting earlier corrective actions and reducing rework. However, the high-dimensional variability inherent to the assembly process introduces considerable uncertainty into the predictive models. Even at early stages as framing, dimensional deviations may already be present, reflecting upstream variations from stamping and sub-assembly operations. Consequently, minor deviations — though still within specification

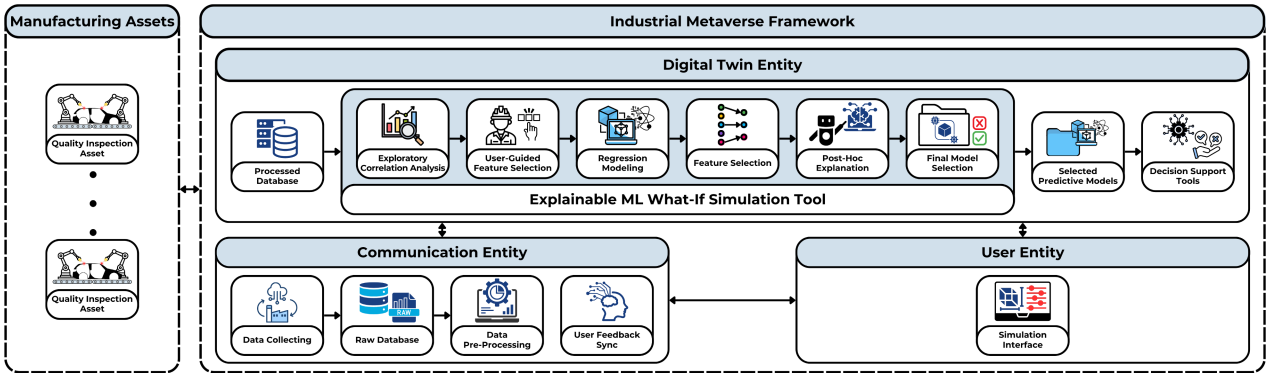

Figure 1. ISO 23247 standard-compliant architecture for exploring correlations through explainable ML-enabled what-if simulation.

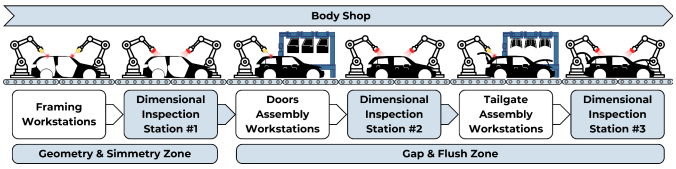

Figure 2. Assembly and inspection stations in the automotive body shop.

limits — can accumulate and propagate across subsequent stages, making it difficult to identify stable and meaningful correlations between measurements at different process stages using purely statistical analysis of historical data. Furthermore, this variability can also obscure the relationships between features measured within the same inspection station, as local deformations caused by the welding process and residual stress often lead to weak or inconsistent correlations.

To address these limitations, the proposed approach integrates explainable ML techniques and exploratory analysis with a what-if simulation tool to enable quality control stakeholders to iteratively explore, interpret, and refine feature correlations even under variability and uncertainty. As a first step, the methodology is applied to the framing stage, where early deviations often emerge. This preliminary validation supports future expansions toward identifying correlations within and across subsequent inspection stages, enabling the development of robust predictive quality processes.

V. Deploying the Explainable ML What-If Tool for Automotive Quality Correlation Analysis

A functional version of the DT architecture depicted in Section III was implemented and deployed to analyze the correlations among quality measurements of the automotive company's framing inspection station. The geometric information acquired by the station was stored in a raw database on the company's internal server and subsequently pre-processed to remove incomplete measurements and discard points considered irrelevant to the quality control process. The processed and structured data are made available to the explainable ML what-if simulation tool, which is integrated with an interactive visual interface supported by the PyQt6 library. The complete workflow of the simulation tool is illustrated in Figure 3.

The analysis process begins with the **simulation scenario configuration**. Users first select a target variable and define filtering criteria to guide the analysis. These criteria include the choice of correlational metrics, the data subsets to be considered (e.g., all measurement points or only critical ones) and the threshold values for correlation coefficients. The tool's what-if capability empowers users to iteratively explore several analytical scenario configurations, facilitating the identification of conditions that may improve predictive model performance.

The tool computes Pearson, Spearman, and Kendall correlation matrices in the **exploratory correlation analysis** module to capture complementary types of relationships. Pearson identifies linear dependencies and is easy to interpret, but is sensitive to outliers and assumes normality. Spearman and Kendall are non-parametric methods that capture monotonic relationships and are better suited for non-linear and non-normally distributed data. Spearman is processing efficient, though less robust to noise, while Kendall offers greater stability under data variability at a higher computational cost. Combining these methods enables cross-validation of correlation patterns in high-variability measurements.

Next, in the **user-guided feature selection** stage, the correlation coefficients are displayed through heatmaps on the interface categorized by threshold intensity: strong (±0.7 to ±1.0), moderate (±0.3 to ±0.7), weak (0 to ±0.3), and custom-defined. These visual cues assist users in selecting input features for the regression models. Domain experts can leverage their knowledge of the production process to discard suspicious variables that, despite showing high statistical correlation, may represent spurious relationships, e.g., when the correlated variables are located in unrelated areas of the vehicle. For non-expert users, the ability to interactively test different feature combinations through the what-if simulation fosters process understanding and supports more informed decision-making.

The **regression modelling** module trains distinct ML models based on the user-selected input features of each correlation matrix. For instance, a Linear Regression is constructed for the Pearson subset, as it is well-suited for capturing linear relationships. A Random Forest model is trained on Spearman-derived features due to its ability to model complex, non-linear interactions and handle high-dimensional data effec-

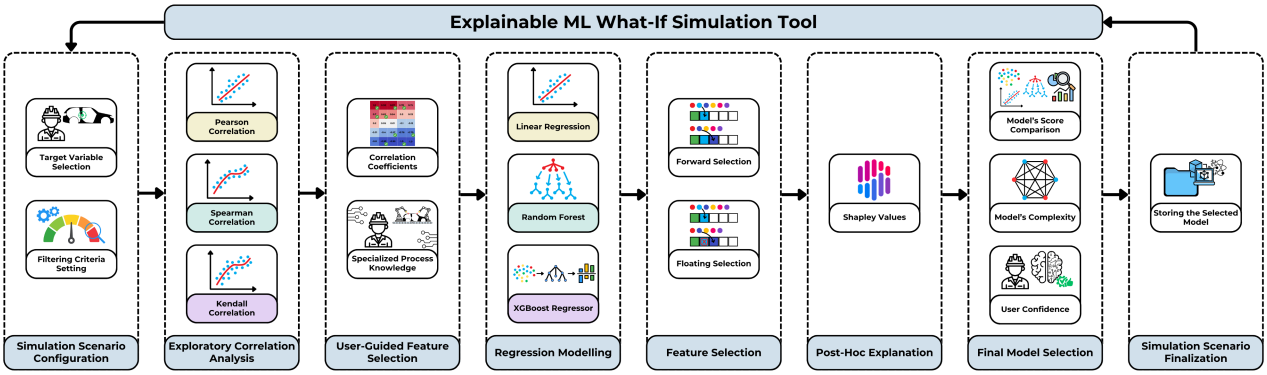

Figure 3. Functional workflow of the explainable ML what-if simulation tool applied to the automotive quality measurement case study.

tively. Features selected via Kendall's analysis pass by an XGBoost Regressor, which offers strong regularization and generalization capabilities, being reasonable for subtle patterns in noisy or uncertain data. The association between correlation types and regression models was established empirically based on the complementary characteristics of each method and their alignment with the nature of the selected features. Each model is independently built and later subjected to comparative evaluation by the user in subsequent modules.

The **feature selection** module enhances the regression model's performance by applying the forward selection algorithm to add features that improve prediction accuracy iteratively and the floating selection algorithm to dynamically remove features that no longer contribute to the model once new features are included. This process helps reduce overfitting and eliminates input variables irrelevant to the prediction of the target variable. It also decreases model complexity, making the models more interpretable. The selected and excluded features are displayed through the interface, helping users to assess their relative importance or their lack.

Once the models are trained, SHAP algorithms are applied for **post-hoc explanation** to quantify the contribution of each input variable to the predicted outcome, offering a model-agnostic interpretability layer. The Shapley values are displayed in the interface through beeswarm plots, illustrating how each measurement point influences predictions across different quality measurement instances. Besides providing experts and non-experts with insights into how structural points affect one another, these explanations allow domain experts to compare the model's reasoning with their knowledge of the production process. This alignment enables them to assess whether the identified relationships reflect real shop floor behaviour, increasing confidence in the model's outputs.

The tool supports multi-criteria decision-making for the **final model selection** among the regression alternatives. Besides the user's confidence, guided by SHAP-based explanations, various performance metrics such as R^2 (coefficient of determination), MAE (Mean Absolute Error), and MSE (Mean Squared Error) are presented, along with each model's complexity, expressed by the number of selected input features. These evaluations allow users to balance predictive strength and practical usability when selecting the most appropriate model. Once selected, the model is stored during the **simulation scenario finalization** and made available for future integration into decision support tools.

The execution of a simulation scenario using the developed interface is shown in Figure 4. The left panel corresponds to the simulation scenario configuration. In the example, a critical point is selected as the target to explore correlations using the three correlation matrices. The correlation threshold is set to "strong", and the user chooses to explore correlations only among the other critical points. In the right panel, some of the variables correlated with the target are displayed as buttons using the Seaborn colour scheme — the closer to dark red, the stronger the positive correlation; the closer to dark blue, the stronger the negative correlation. The green button represents a variable the user selected for the regression models.

Figure 4. View of the simulation tool interface: simulation scenario configuration (left panel) and user-guided feature selection (right panel). Note: Point names were changed to preserve company data privacy.

In Figure 5, the interface displays the performance metrics of the regression models based on the user-selected input features, later refined through feature selection. In the first panel, "Point_25" is added via forward selection and then removed by floating selection after "Point_03" is included, improving the model's accuracy. The final metrics are followed by post-hoc explanations using SHAP values, shown through beeswarm plots. In the first plot, all selected features have a similar influence; in the second, one feature has a slightly

greater impact; and in the third, a single feature dominates the prediction. Based on the metrics and interpretability, the Random Forest model is selected in this scenario.

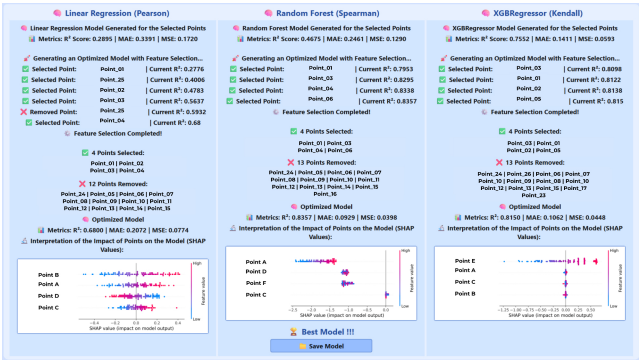

Figure 5. The interface shows the feature selection process, a post-hoc explanation with Shap, and the metrics of the regression models.

VI. Experimental Validation and Discussion

Dimensional inspection data collected over 16 months of production were used for the preliminary validation of the implemented simulation tool. The dataset was pre-processed and divided into two subsets: a training set containing measurements from over 150,000 vehicles and a test set with data from more than 50,000 vehicles. Each vehicle inspection at the framing station includes approximately 205 dimensional measurement points, of which 66 are classified as critical due to their significant impact on the vehicle's structural integrity.

A series of 66 simulations were conducted to evaluate the tool's explainability capabilities, one for each critical point set as the prediction target. Figure 6 shows the average number of correlated features identified for each correlation matrix (Pearson, Spearman, and Kendall), grouped by strength thresholds (strong, moderate, and weak), considering relationships between each critical point and all other measurements. The variation in the number of identified features across matrices and thresholds highlights the diversity of dependency types captured. On average, Pearson yielded the highest number of strongly correlated features, followed by Spearman and Kendall. As expected, a larger concentration of features falls into the weak correlation range. However, a high statistical correlation does not necessarily imply a causal relationship, underscoring the importance of combining statistical insights with explainable mechanisms to validate variable relevance.

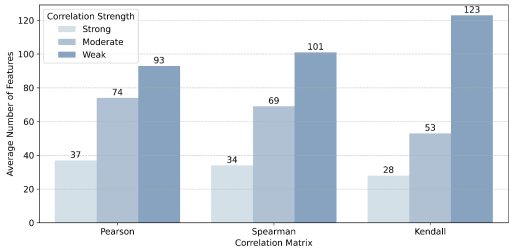

Figure 6. Average number of features per correlation strength and type.

A non-expert user scenario was considered for the user-guided feature selection, where all features with strong or moderate correlation with the target were initially selected. All three regression models were trained and submitted to the feature selection. Figure 6 presents the average values of R^2, MAE, MSE, and the number of features before and after the feature filtering process, covering the 198 models generated (three per target variable) and the 66 final models selected by the user (one per simulation). The results refer to model evaluation using the test dataset. In all cases, a substantial increase in R^2 and a reduction in both MAE and MSE were observed after applying the forward and floating selection algorithms, demonstrating how these methods effectively eliminate irrelevant or misleading features and contribute to improving explainability.

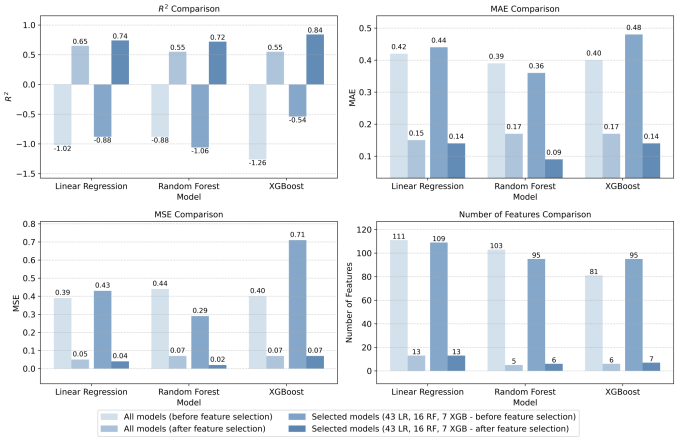

Figure 7. Comparison of average metrics and number of input features before and after feature selection, across all generated and selected regression models.

Among the selected models, XGBoost achieved the highest average R^2, Random Forest reached the lowest MAE and MSE values, and Linear Regression was selected most frequently (43 out of 66 cases), reinforcing the value of a mixed-model strategy. The tool successfully produced accurate and explainable predictive models even under a non-expert user scenario. Future validation by domain experts may further demonstrate the added value of expert feedback in improving model quality and identifying potentially spurious relationships.

Aside from supporting model selection, the SHAP values were used to identify the 10 most influential measurement points in the structural quality of the vehicle during the framing process. Figure 8 presents how often each point appeared among the five features with the highest impact on model predictions. The results show a balanced distribution between critical and non-critical points. While critical points confirm its strong dependencies in predicting structural deviations, including non-critical points highlights the tool's ability to uncover subtle or unexpected correlations. These insights can help specialists detect new patterns of defect propagation and refine quality control strategies.

SHAP value distributions were analyzed to assess the uncertainty associated with the predictions of the selected models.

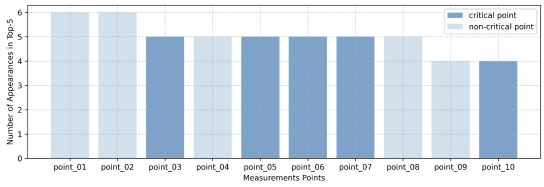

Figure 8. Most frequently occurring features among the top-5 SHAP impacts across all selected models.

Each model was applied to a randomly sampled subset of 1,000 test instances. Figure 9 presents the distribution of average standard deviations for the SHAP values of input features, indicating the variability of feature contributions across samples. Most models concentrate below a 0.2 threshold, suggesting consistent feature importance and stable model behaviour. The models with higher deviations may signal sensitivity to input variation or potential instability. These results support an uncertainty-aware interpretability approach, increasing confidence in model predictions and helping identify models that may require retraining or expert review before deployment in quality control routines.

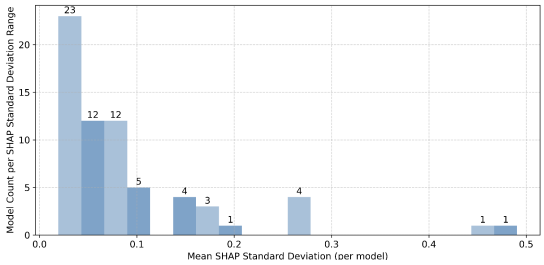

Figure 9. Uncertainty-aware analysis through the distribution of the mean SHAP standard deviation across the selected models.

VII. Conclusion and Future Works

This paper has introduced a what-if simulation platform that combines exploratory correlation analysis and explainable ML to support dimensional quality assessment in the automotive industry. The platform enables users to interactively investigate correlations and build interpretable predictive models through an ISO 23247-compliant DT framework. The experimental validation on real inspection data demonstrated the platform's effectiveness in identifying relevant features, improving model accuracy, and enhancing transparency through post-hoc SHAP explanations. The results reinforce the value of combining statistical correlations with explainable ML and expert-in-the-loop strategies. The feature selection approach applying forward and floating methods substantially reduced the model complexity while increasing accuracy, while the SHAP-based analysis enabled understanding of feature relevance and offered a layer of uncertainty-aware interpretation that enhances trust in the predictions.

Future work will extend the methodology across other inspection stages, incorporating multistage correlations and evaluating the platform with domain experts in the production environment. Additional efforts will focus on integrating the predictive models to develop predictive and decision-supporting tools for the quality process and explore more robust strategies for model uncertainty quantification and improvement.

Acknowledgments

This work was partially supported by the HORIZON-CL4-2021-TWIN-TRANSITION-01 openZDM project under Grant Agreement No. 101058673 and was also supported by national funds through FCT/MCTES (PIDDAC): CeDRI, UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and UIDP/05757/2020 DOI: 10.54499/UIDP/05757/2020); and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020). The author Alexandre O. Júnior thanks FCT, Portugal, for the PhD grant BD/03967/2023 (DOI: 10.54499/2023.03967.BD).

References

[1] A. Gallala *et al.*, "Digital Twin for human–robot interactions by means of Industry 4.0 Enabling Technologies," *Sensors*, vol. 22, no. 13, p. 4950, 2022.

[2] J. Guo *et al.*, "Industrial metaverse towards Industry 5.0: Connotation, architecture, enablers, and challenges," *Journal of Manufacturing Systems*, vol. 76, pp. 25–42, 2024.

[3] A. O. Júnior *et al.*, "Simulation on Digital Twin: Role of Artificial Intelligence and Emergence of Industrial Metaverse," in *2024 IEEE 33rd International Symposium on Industrial Electronics (ISIE)*, 2024, pp. 1–6.

[4] H. Zhang *et al.*, "Artificial Intelligence-Enhanced Digital Twin Systems Engineering Towards the Industrial Metaverse in the Era of Industry 5.0," *Chinese Journal of Mechanical Engineering*, vol. 38, no. 1, p. 40, 2025.

[5] M. K. Msakni *et al.*, "Using machine learning prediction models for quality control: a case study from the automotive industry," *Computational management science*, vol. 20, no. 1, p. 14, 2023.

[6] K. Kobayashi and S. B. Alam, "Explainable, interpretable, and trustworthy AI for an intelligent digital twin: A case study on remaining useful life," *Engineering Applications of Artificial Intelligence*, vol. 129, p. 107620, 2024.

[7] A. Bujari *et al.*, "A Layered Architecture Enabling Metaverse Applications in Smart Manufacturing Environments," in *2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom)*, 2023, pp. 585–592.

[8] T. Kreuzer *et al.*, "AI Explainability Methods in Digital Twins: A Model and a Use Case," in *International Conference on Enterprise Design, Operations, and Computing*. Springer, 2024, pp. 3–20.

[9] J. Cohen and X. Huan, "Uncertainty-aware explainable AI as a foundational paradigm for digital twins," *Frontiers in Mechanical Engineering*, vol. 9, p. 1329146, 2024.

[10] M. Rabiee *et al.*, "Towards explainable artificial intelligence through expert-augmented supervised feature selection," *Decision Support Systems*, vol. 181, p. 114214, 2024.

[11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[12] J. Hong *et al.*, "Enhancing the product quality of the injection process using explainable artificial intelligence," *Processes*, vol. 13, no. 3, p. 912, 2025.

[13] S. Suhail *et al.*, "Enigma: An explainable digital twin security solution for cyber–physical systems," *Computers in Industry*, vol. 151, p. 103961, 2023.

[14] R. S. Peres *et al.*, "Multistage quality control using machine learning in the automotive industry," *IEEE Access*, vol. 7, pp. 79 908–79 916, 2019.

[15] E. Birihanu and I. Lendák, "Explainable correlation-based anomaly detection for industrial control systems," *Frontiers in Artificial Intelligence*, vol. 7, p. 1508821, 2025.

[16] *Automation systems and integration – Digital twin framework for manufacturing – Part 2: Reference architecture*, International Organization for Standardization Std. ISO 23 247-2:2021, 2021.