

OPTIMIZING PAINTING OPERATION IN AN AUTOMOBILE PAINTSHOP USING ML APPROACH

Isaac O. Olalere*, and Kemllal Ramdass

Department of Industrial Engineering, University of South Africa, UNISA
Florida Campus, Roodepoort, Johannesburg, South Africa

*Corresponding email: opeflan@gmail.com

Article history

Received
1st July 2025
Revised
25th March 2026
Accepted
25th March 2026
Published
2nd June 2026

ABSTRACT

Optimizing paint operations is crucial to ensure product quality, consistency, and customer satisfaction in the automobile paintshop. Detailed analyses and understanding of the paint process provide a comprehensive optimization approach to increase productivity and reduce operational costs while maintaining a competitive advantage. The study aims to improve the painting operation in an automobile paintshop through key painting parameters, implementing quantitative approach techniques, using both random forest regression and reinforcement ML model, and evaluating the effectiveness of the optimized model for enhancing the painting processes. The painting defects and the causes were analysed using a cause-and-effect diagram, and the corresponding parameters associated with the defects were used for data modelling to optimize the process. Random forest regression model predicts the good products, number of scraps, and number of reworks with an accuracy of 70.6%, while the Q-learning model has a prediction accuracy of 94.6%. The model was validated by adjusting the system parameters to the predicted model data, which indicates a significant improvement in the FTT, number of scraps, and reworks. The significance of the study is its potential to drive operational and process improvements, to increase the organization's productivity by reducing paint defects.

Keywords: *Paintshop; random forest regression; reinforcement learning; First Time Through; Defects*

© 2026 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

An original equipment manufacturer (OEM) whose specialty is the production of paints, solvents, and tints, as well as a variety of surface-finish products, plays a crucial role in the automobile manufacturing sector. The diverse techniques of paint application are crucial in achieving desired coating properties and ensuring optimal performance of the manufacturing process. According to Akafuah, Poozesh [1], brushing, rolling, and spraying are some of the methods of applying paint coatings onto surfaces. As described by Eley [2], the use of brushes when applying paint is known as brushing, and it is more precise and offers more control over the thickness of the coat. The challenges with brushing, however, it tends to introduce brush marks and variation in the evenness of the coat thickness. On the other hand, spraying, as stated by Brega, Trifonov, and Voronko

[3], utilizes compressed air or airless systems to atomize paint into fine droplets and deposit them onto surfaces, yielding uniform coverage and smooth finishes. Spraying is preferred for large-scale projects due to its speed and efficiency, but it requires careful masking and ventilation to control overspray. Additionally, rolling, as explained by Maile [4], is the process of applying coatings onto surfaces using paint rollers. This procedure is more efficient, yields improved quality, and is highly applicable to flat and rough surfaces. Several factors determine the choice of paint application techniques and methods, which offer significant advantages depending on the type of surface, the desired surface finish, and the coating thickness.

1.1 Study Background

Advancements in the painting technology process have resulted in the growing adoption of adaptable systems and applications, such as robots, automated technology, and precision manufacturing in the painting industry [5]. These advancements lead to improvements in the painting process through a precise coating process, resulting in a consistent thickness of the coating. It is, however, important to prepare the surface of the painted part thoroughly to ensure a quality finish that is durable with good adhesion [6]. Coating operation is highly essential in the automobile paintshop because it adds a layer of protection and beauty to the painted part, increasing the resistance to wear, corrosion, and harsh weather conditions.

Considering the overwhelming market potential and coating footprint, well over 130 countries, these tasks and adventures have gained many customers upfront, both within several industries [7].

The automobile company used for this study is a merger of two industries, with one of the companies being a major coating performance company since 2013, and the other a major coating manufacturer established in the year 1889 in South Africa. The company prides itself on a range of coating solutions that offer standard and distinctive quality coatings to its customers. Considering the growing global market competition, the company pushes to capitalize on emerging opportunities together with continuous improvement approaches to stay abreast of competition, focusing on satisfying customers' needs.

1.2 Problem Statement

Despite a global footprint spanning 130 countries and territories, serving a broad customer base across multiple industries, including automotive, transportation, aerospace, marine, architectural, and industrial sectors, the processes are not sufficiently optimized. This is due to variations in process parameters, equipment maintenance, and unstandardized procedures. According to Bonaduce, Duce [8], the transformative nature of paint, various application methods, raw materials, surface preparations, and environmental conditions are factors that play an important role in the efficient painting process. These deficiencies lead to increased waste, rework, and higher operational costs. This study aims to identify production inefficiencies within the coating and paint shop of XYZ company and optimize the processes through data modelling and analysis using an ML approach to enhance operational efficiency, product quality, and cost reduction. A data-driven approach using AI models would offer improvement to the overall paint processes using trained ML models.

2.0 MATERIALS AND METHODS

There are several advantages associated with adopting a data-driven approach for improved maintenance costs, reducing the frequency of breakdowns, and increasing production output in a manufacturing environment. This implies that the manufacturing facilities have higher availability, with a corresponding reduced downtime, increased spares inventory, and production output, which ultimately increases the profit margin of the company. The coating industry has recently witnessed a growing demand for coating services, leading to these industries optimizing the coating and painting processes in their operations [9]. Automated painting systems equipped with robotic arms and advanced sensors offer precise control over paint deposition, resulting in higher consistency and uniformity in coating thickness [10].

2.1 Review of Literature on Paint Application and Process Efficiency

A broad spectrum of cutting-edge trends and technologies is emerging in the field of paint application and process improvement. Some recent studies have focused on the treatment of wastewater produced during the painting process due to the different pollutants, encompassing organic compounds, suspended solids, heavy metals, and coloring agents contained in the water [11, 12]. Other studies have explored the advancement in Computer-aided design (CAD) technology in the paint and coat manufacturing [13], while other studies have focused on optimizing energy requirements associated with producing paints and coatings [14]. Smart coatings and nanotechnology are playing an increasingly important role in the evolution of paint application. This includes Intelligent Superhydrophobic Coating (ISC), which focuses on the development and engineering of surface micro/nanostructures aimed at forming water-repellent spherical droplets, thereby providing self-cleaning, anti-pollution, and related capabilities [15, 16]. Smart coatings embedded with responsive materials demonstrate self-repairing capabilities, resist corrosion, and improve durability, ultimately prolonging the life of coated surfaces [17].

Furthermore, through real-time monitoring and analysis, digitalization and data analytics are reshaping paint application and process improvement. IoT devices and advanced sensors track key variables such as temperature, viscosity, and flow, allowing for predictive maintenance and the strategic improvement of operational parameters [18]. Other studies have indicated that automotive paint curing process parameters also affect the efficiency of the overall painting process. Pendar, Cândido [19] performed a key adjustment in the volume and direction of heated air supplied to the intake using a conjugate heat transfer (CHT) algorithm, providing a reliable operational condition and enabling optimal oven performance. With an impressive 98% accuracy, the model provided a reliable tool for determining the optimal operational conditions for the paint curing process. In another study by Corbu, Nitescu, and Edelhauser [20], achieving minimal tolerances, the color-identification process was enhanced through the optimal calibration of correction factors and the analysis of influencing parameters. Samples of the dataset were taken and used to evaluate the performance of two color-distinction formulas using visual testing methods for identifying colour differences. This approach reduces painting rework and the number of scrap parts due to colour differences. Guettler, Sandgren [21], on the other hand, proposed a self-programming painting cell, using 3D-scanning, multi-physics painting simulations using terahertz technology enabled through a cyber-physical system. The approach was aimed at improving and ensuring atomizing quality during painting processes for small lot sizes of automotive components. Increasing focus on sustainability has driven the advancement of environmentally friendly paint technologies. Water-based coatings, bio-based resins, and low-VOC formulations are emerging as viable and eco-conscious alternatives to conventional solvent-based paints [22]. To accurately determine the most viable solutions, a multi-objective particle

swarm optimization (MOPSO) algorithm has been developed, integrating targeted improvement strategies specific to the problem domain on environmental emissions during the painting process [23]. The extensive experiments showed that the adopted MOPSO optimization model was capable of significantly reducing the emissions produced during the auto-painting process. Other studies have explored optimizing energy consumption in the automotive paint industry. Wang, Li, and Huang [24] optimized scheduling by choosing effective batch and sequencing policies, with the goal to enhance the quality of paint application and streamline production operations, reducing rework. The power and material utilization in the painting operation is reduced. Reducing the number of scrapped painted parts due to painting deformities and the number of parts repainted is essential in increasing and optimizing the painting process in the automotive industry. The process parameters that directly affect the quality output of the painting process are determined and accurately measured during operation. These parameters are essential to the painting process to prevent rework and optimize operations. These conditions are closely monitored during manufacturing using advanced sensors and a process-monitoring system.

2.1.1 Temperature

Temperature represents the average kinetic energy of particles in a substance, commonly measured in Celsius (°C), Kelvin (K), or Fahrenheit (°F) according to Howells [25]. According to Cavalcante, Vasconcelos [26], instantaneous temperature fluctuation caused by the changing weather conditions makes it imperative to perform painting operations in a temperature-controlled environment. This has a substantial effect on paint applications, affecting drying times, curing procedures, and the final properties of the coating [27]. The temperature can be determined using Equation 1.

$$T=Q/mc \quad 1$$

Where the heat energy is Q, the mass is m, and the heat capacity, c.
With the other variables known, the system temperature can be determined. According to Müller and Schackmann [28], radiation-curing coatings can affect the properties of the paint on the surface area it's applied to.

2.1.2 Viscosity

According to Anderson, Bruno, and Smith [29], viscosity is a measure of a fluid's resistance to deformation at a set rate and is often referred to as its 'thickness' or internal friction. This property, which governs how fluid flows, significantly impacts the quality and finish of paint applications. It is a key factor in painting application, influencing both the quality of the finish and the paint's ability to atomize and spread evenly during the spraying process [30]. Viscosity can be expressed as illustrated in Equation 2.

$$\eta = \tau / \dot{\gamma} \quad 2$$

Where η is the viscosity,
 τ is the shear stress, and $\dot{\gamma}$ is the shear rate.

Paint that is too viscous can be difficult to spray properly, leading to uneven coverage, orange peel texture, and possible sagging or dripping [31]. If the viscosity is not adjusted, the paint's final finish may exhibit an uneven texture, showing visible brush marks or streaks. This reduces both the aesthetic quality and the protective properties of the coating. However, when the viscosity is within the ideal range, it allows for proper atomization,

leading to smooth and consistent coverage and a flawless finish [32]. However, paint with low viscosity is thinner and more prone to over-atomization, which can cause problems like excessive overspray, reduced coverage, and the risk of runs or sags in the applied layer [28]. The quality effect results in a thinner layer of paint, which can cause poor coverage and a reduction in the coating's protective properties.

2.1.3 Pressure

The pressure exerted during the paint spraying process is a key factor that significantly influences the quality of the coating's application to the finished product. The pressure settings determine the force with which paint is atomized and applied to the surface. According to Rajagopal [33], painting pressure is the continuous physical force exerted on or against an object by something in contact with it. High pressure can cause over-atomization of the paint, resulting in excessive overspray and possible material waste.

The result is potential material waste and the occurrence of a dry spray effect, where paint particles do not sufficiently wet the surface, leading to a rough, uneven finish [34], and may also result in a rough, sandy texture with inadequate adhesion, which can contribute to early coating failure [35]. Setting the pressure to optimal levels ensures proper atomization of the paint and enhances transfer efficiency during application. This leads to an even application of paint, with minimal overspray, resulting in a smooth, high-quality finish [36]. This indicates that the final product will have a seamless, uniform surface, offering excellent coverage and durability, which enhances both the aesthetic value and protective qualities of the coating.

Therefore, low pressure causes incomplete atomization, which results in larger paint droplets being deposited. This can lead to a blotchy surface, weak adhesion, and an uneven finish, with the possibility of runs and sags in the coating. [37, 38], hence, it is important to monitor it during the painting process.

2.2 Optimization Techniques

Optimization is the process of adjusting decision variables or process parameters to determine the optimal solution that meets the constraints imposed by the system [39]. A study by Chen, X., et al. [15] adopted an optimization technique on critical parameters such as path interval, spray tool orientation, and the motion velocity of the spray tool to create a model that improves coating thickness based on the path parameter optimization. This approach improves the painting efficiency, reduces rework, and sustains the quality of the paint layer on the product surface. A similar but different approach was adopted by Zunaidi [40] in a study that examines the painting of aircraft parts using a gantry robot with an automated spray-painting gun, specifically designed to control the spray path. The study concluded that integrating the manual process with an automated painting system resulted in a 30-40% increase in speed, a reduction in setup time, and an increase in the painting booth's operational capacity. A different optimisation approach was adopted in another study by Pendar, Cândido, and Páscoa [41] using Gradient Gauss linear regression method, to reduce the time of paint application and the amount used up. The study optimized the spray plume pattern to achieve a smaller diameter, while a machine learning model (Gradient Gauss linear regression) accurately predicted paint flow, thereby minimizing both computational time and associated costs.

This study optimizes the painting process in an automobile paintshop by creating a machine learning model that is based on the process parameters to predict the optimal parameters level that yield the maximum productivity based on historical data. Unlike most paintshop optimization approaches that focus on spraying nozzles and tools, and precision in tool handling, this study focuses on process parameters associated with the spraying process to reduce or eliminate rework. A study by Pendar, Candido [42] applied

a Machine Learning algorithm, using a variational auto-encoder and artificial neural network (ANN) model, effectively encodes temperature and velocity fields for curing the painting process while Kwofie, Perera [43] analyzed carried out alternate least squares (ALS) reconstructions of the infrared (IR) spectra for painting nonconformities in automobile vehicles. These ML approaches, however, build the optimisation model from the measured parameters from the finished product, unlike the novel approach in this study that develops the model from the process parameters that directly influence and determine the product quality during the painting process. Adopting this approach also implies that the number of reworks can be reduced. In addition, the Q-learning algorithm was adopted due to its application in solving complex classification problems, such as this study, determining the

2.3 Methodology

This study identifies and monitors key parameters during the painting process to establish the most suitable operating conditions. According to Torrent Solà [44], temperature, viscosity, and pressure are essential parameters that determine the production First Time Through (FTT) in an automotive painting process. The research study measures both the process data during the painting process and the production KPIs, such as daily total units painted, rework and scrap, and efficiency. By using this integrated approach, the narrative becomes more coherent, allowing for a clearer link between the data, its meaning, and the surrounding context. The collected data is used to develop an ML model that determines the parameter level that produces an optimal result. The schematic of the research design is illustrated in Figure 1.



Figure 1: The Pareto Chart of the Paintshop Defects

The system parameters are measured every one-hour interval, and the total number of units painted, the reworks and scrap, efficiency measured in FTT, and daily production is also recorded. The temperature of the spray booth is measured hourly using the temperature sensor, and the pressure is monitored and read by the pressure gauges located downstream of the pump transferring paint from the tank, or those on the compressor supplying air to the spray booth. These data are cleaned, and some features are extracted. Data from the painting production line was collected, with the resulting daily production output. The data is then analysed, and the appropriate ML model to establish ideal production settings is employed.

Two separate models are developed for optimizing the painting process. The first Model is a regression-based optimization model that predicts the optimal level of the parameters that give the best painting results, while the second model is a deep learning model that treats the problem as a decision-making process using a reinforcement learning (RL) algorithm. Random forest (RF) regression algorithm is used for its major advantages over other regression algorithms. It is notably effective in scenarios with few samples, unlike several other algorithms that tend to face challenges due to insufficient data, and it is appropriate for nonlinear or approximately linear problems [45, 46]. Deep

reinforcement learning algorithms are capable of autonomously extracting concise, low-dimensional representations from high-dimensional data such as images, text, and audio [47, 48], which is a novel advantage that the algorithm offers in this study. Since there are arrays of process parameters and their corresponding variations in the product quality, reinforcement learning using the Q-learning algorithm helps to determine the combination of parameters that yields the optimal, least number of scraps and reworks. The study implements the two ML approaches to compare the models' results and determine the model with the least performance error.

2.3.1 Random Forest Model Development

A decision tree regressor splits the feature space recursively by minimizing an impurity measure (e.g., Mean Squared Error, MSE), and the prediction at each node is the average of the target variable values in that region. For a given dataset $D = \{(x_i, y_i)\}_{i=1}^n$ is the feature vector and y_i is the target variable, and the tree partitions the space by choosing a feature j as a threshold s that minimizes the given in Equation 3.

$$MSE = \sum_{i \in L} (y_i - \bar{y}_L)^2 + \sum_{i \in R} (y_i - \bar{y}_R)^2 \quad 3.$$

Where L and R are the left and right child nodes once split, while \bar{y}_L and \bar{y}_R are the corresponding mean target values, respectively.

Random Forest Regression, however, improves a single decision tree by aggregating multiple decision trees. It does this by bootstrap sampling, feature randomization, tree training, and prediction aggregation that averages the predictions of all trees.

For M trees, the final prediction for a given input x is given in Equation 4.

$$\hat{y} = \frac{1}{M} \sum_{m=1}^M T_m(x) \quad 4$$

Where $T_m(x)$ is the prediction from the m -th tree.

The optimization aims to determine the optimal conditions. $x^* = (x_1^*, x_2^*, \dots, x_d^*)$ that maximize the number of good products while minimizing scrap and rework.

The objective function is formulated as given in Equation 5.

$$\max_x f(x) = \hat{y}_{good}(x) - (\hat{y}_{scrap}(x) + \hat{y}_{rework}(x)) \quad 5$$

Where $\hat{y}_{good}(x)$ is the predicted number of good products, $\hat{y}_{scrap}(x)$ is the predicted number of scrap products, and $\hat{y}_{rework}(x)$ is the predicted number of rework products. Thus, the function to minimize (since optimizers often work on minimization problems) is given in Equation 6.

$$\min_x - \left(\hat{y}_{good}(x) - (\hat{y}_{scrap}(x) + \hat{y}_{rework}(x)) \right) \quad 6$$

Each feature is constrained within its observed minimum and maximum values represented in Equation 7.

$$x_j^{min} \leq x_j \leq x_j^{max}, \forall j = 1, \dots, d \quad 7$$

Where x_j^{min} and x_j^{max} are the minimum and maximum observed values of the feature j.

Optimization Algorithm

As a quasi-Newton approach, the L-BFGS-B algorithm relies on an approximation of the Hessian matrix to optimize the function efficiently while respecting bounds. The update step is given in Equation 8.

$$x^{(k+1)} = x^{(k)} - \alpha H^{-1} \nabla F(x^{(k)}) \quad 8$$

Where H approximates the Hessian matrix, $\nabla F(x^{(k)})$ is the gradient of the objective function. The algorithm iterates until the convergence criteria (e.g., small gradient norm) are met.

2.3.2 Reinforcement Learning (using Q-learning) Model Development

The paint shop conditions are modeled using the Markov Decision Process (MDP) characterized by the set of states, which are the combinations of discretized feature values such as temperature, pressure, and viscosity, set of actions for each feature, which are to increase, decrease, or remain unchanged. To determine the reward function R, it is evaluated based on regression models as indicated in Equation 9.

$$R(s, a) = \text{Good Products} - (\text{Scrap} + \text{Rework}) \quad 9$$

This implies that a transition to the next state, s', would result from applying action a to state, s.

Using the Q-Function, $Q(s, a)$ reflects the anticipated total reward from taking action a in state s and proceeding with the best possible strategy. The optimal Q-value function satisfies the Bellman Equation 10.

$$Q^* = E \left[R(s, a) + \gamma \max_{a'} Q^*(s', a') \right] \quad 10$$

Where R(s, a) is the instant reward, $\gamma \in [0, 1]$ is the discount factor, and the next state and actions are s', and a' respectively.

The Q-Learning algorithm approximates the optimal Q-function using the update rule as illustrated in Equation 11.

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[R(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad 11$$

Where $\alpha \in (0, 1]$ is the learning rate, the term in the bracket is the difference error. The reward function from the regression model can be mapped to real-value features using bin centers (\bar{x}_i), for a given state, s. The model's prediction can be expressed as;

Good(s), Scrap(s), Rework(s) = Predictions from trained Random Forest models

$$R(s, a) = \text{Good}(s) - [\text{Scrap}(s) + \text{Rework}(s)] \quad 12$$

Equation 12 is the optimization objective that maximizes R, which corresponds to maximizing production quality. After sufficient iterations, the Q-table Q(s, a) contains

estimates of the long-term value of actions. The optimal state s^* , is the state with the highest maximum Q-value as expressed in Equation 13.

$$s^* = \arg \max_s \max_a Q(s, a). \quad 13$$

Therefore, the corresponding decoded real-valued feature vector gives the optimal process conditions for maximizing good products and minimizing defects and scraps.

2.3.3 Data Structure and Model Development

The paint temperature, pressure, humidity, and viscosity are the process input parameters for the model development, while the output parameters are the number of scraps, FFT, and reworks. Daily production targets of 720 units are painted, while system parameters are captured hourly, and the number of rework and scrap is recorded per shift. The booth temperature data measurement for the month of July is illustrated in Table 1, with a similar structure to the humidity, pressure, and viscosity.

Table 1: Booth Temperature for July

		Temperature (°C)																	Specification (23 ± 2°C)						
Month:	July																								
Date:	Time																								
	06h00	07h00	08h00	09h00	10h00	11h00	12h00	13h00	14h00	15h00	16h00	17h00	18h00	19h00	20h00	21h00	22h00	23h00	00h00	01h00	02h00	03h00	04h00	05h00	
1	22	23	24	26	23	25	23	28	22	23	23	23	23	25	24	26	24	23	24	25	23	23	24	22	
2	22	23	25	26	22	22	25	24	25	25	23	26	23	25	25	23	23	26	25	24	24	23	23	23	
3	23	23	28	23	25	24	24	24	28	23	23	23	23	23	25	25	23	25	25	24	25	23	23	23	
4	22	21	22	22	22	23	23	23	24	25	27	23	23	25	24	24	24	25	25	25	24	21	21	21	
5	21	22	24	23	22	24	24	24	25	25	26	25	24	24	22	23	24	23	23	23	22	22	22	22	
6																									
7																									
8	23	23	28	23	25	24	24	24	28	23	23	23	23	23	25	25	23	25	25	24	25	23	23	23	
9	22	23	25	26	22	22	25	24	25	25	23	26	23	25	25	23	23	26	25	24	24	23	23	23	
10	22	21	22	22	22	23	23	23	24	25	27	23	23	25	24	24	24	25	25	25	24	21	21	21	
11	21	22	24	23	22	24	24	24	24	25	25	26	25	24	24	22	23	24	23	23	23	22	22	22	
12	22	23	24	26	23	25	23	28	22	23	23	23	23	25	24	26	24	23	24	25	23	23	24	22	
13																									
14																									
15	22	23	24	22	29	25	23	23	22	23	23	23	23	25	27	25	24	23	24	25	23	23	24	22	
16	21	22	24	23	22	24	24	24	25	25	26	25	24	24	22	23	24	23	23	23	22	22	22	22	
17	22	23	25	25	26	22	25	24	25	25	23	25	23	25	25	23	23	28	25	24	24	23	23	23	
18	22	21	22	29	22	23	23	23	24	25	23	23	23	25	24	24	28	25	25	25	24	21	21	21	
19	23	23	27	23	25	24	24	24	22	23	23	23	28	23	25	25	23	25	25	24	25	26	23	23	
20																									
21																									
22	23	23	25	26	22	22	25	24	25	25	23	26	23	25	25	23	23	26	25	24	24	23	23	23	
23	22	21	22	22	22	23	23	23	24	25	27	23	23	25	24	24	24	25	25	25	24	21	21	21	
24	22	22	24	23	22	24	24	24	25	25	26	25	24	24	22	23	24	23	23	23	22	22	22	22	
25	23	23	25	25	26	22	25	24	25	25	23	25	23	25	25	23	23	28	25	24	24	23	23	23	
26	22	22	24	23	22	24	24	24	25	25	26	25	24	24	22	23	24	23	23	23	22	22	22	22	
27																									
28																									
29	24	21	22	22	22	23	23	23	24	25	28	23	23	25	24	24	24	25	25	25	24	22	22	21	
30	22	22	24	23	22	24	24	27	25	25	22	25	24	25	22	23	24	23	23	23	23	23	22	22	
31	23	23	25	25	26	22	25	24	25	25	23	25	23	25	25	23	26	24	25	24	24	23	23	23	

The input data for the training set is determined by evaluating the daily mean, median, variance, and standard deviation of each parameter and corresponding FFT, and the number of rework and scrap is recorded as the output variable. The dataset with 96 entries is loaded from a CSV file and preprocessed by removing outliers in the FFT variable (values > 750) to improve model stability and prediction accuracy.

Input features included process variables (e.g., temperature, pressure, viscosity, and humidity), while target variables were FFT (performance metric), Scrap, and Rework.

Random Forest was chosen due to its ability to capture nonlinear relationships and handle interactions between process variables.

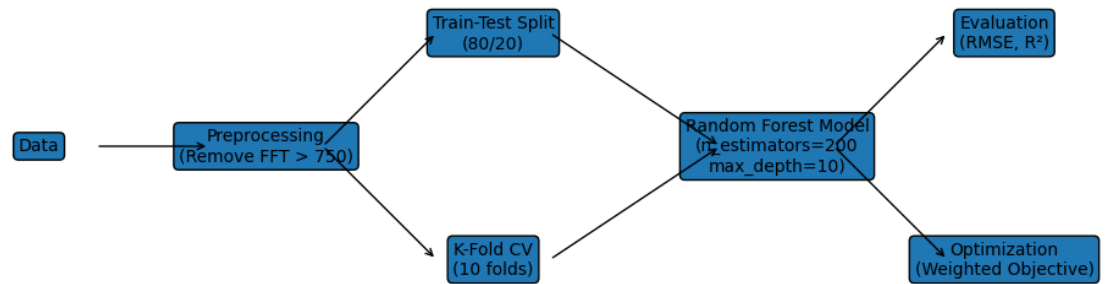


Figure 2 : Model Configuration and Workflow

The model configuration is illustrated in Figure 2 with the number of trees being 200, the maximum tree depth of 10, and the fixed random seed being 42. The data is split into 80% training set and 20% test set, and a random seed of 42 is used to ensure reproducibility. The data is shuffled, and 10-fold cross-validation is applied to ensure robust estimation of model performance, especially given the relatively small dataset. The model performance is evaluated using Root Mean Squared Error (RMSE) and the coefficient of determination (R^2) across folds.

A manual hyperparameter tuning approach is used to test different values of $n_estimators$ and max_depth and to select parameters that provide low cross-validation RMSE, stable performance (low standard deviation across folds), and good generalization (similar CV and test results). The selected final parameters (200 trees, depth = 10) are based on this balance, and the final models are trained on the entire dataset to maximize learning before optimization. A multi-objective optimization is formulated to maximize FFT, and reduce both scrap and rework, and this is performed using *scipy.optimize.minimize* with bounds defined by observed feature ranges as illustrated in equation 14.

$$Objective = -FFT + Scrap + Rework \quad 14$$

The implementation is carried out in Python, utilizing key libraries such as scikit-learn for model development and validation, pandas and NumPy for data processing, and SciPy for optimization. Visualization is performed using Matplotlib and Seaborn.

3.0 RESULTS ANALYSIS

The paintshop production data comprises the daily numbers of rework, scrap, and FFT as illustrated in Table 1. The data for June 2024 is presented in Table 2, with the paint production plan being 14,400 units and a production target of 98% of the plan production. The actual production was 95.03%, which is 4.97% short of the target, either due to repaints (rework) or numbers scrapped.

Table 2: Production Data for June 2024

Date	Units Painted (Daily Target)	Daily FFT	Daily Rework	Daily Scrap	Daily FFT%
1	No production	0	0	0	0
2	No production	0	0	0	0

3	720	680	38	2	94.44
4	720	640	80	0	88.89
5	720	665	52	3	92.36
6	720	660	59	1	91.67
7	720	700	19	1	97.22
8					
9					
10	720	600	115	5	83.33
11	720	650	67	3	90.28
12	720	700	18	2	97.22
13	720	680	40	0	94.44
14	720	630	90	0	87.50
15					
16					
17	720	720	-1	1	100
18	720	711	6	3	98.75
19	720	701	18	1	97.36
20	720	709	11	0	98.47
21	720	699	21	0	97.08
22					
23					
24	720	701	18	1	97.36
25	720	715	5	0	99.31
26	720	705	15	0	97.92
27	720	712	8	0	98.89
28	720	706	13	1	98.06
29					
30					
Total	14400	13684	24	692	

The paintshop defects causing high reworks and scraps were identified and recorded, and the top 5 defects are illustrated in Table 3. The defects indicated that three of the five top issues were caused by process parameters and performance operations during the painting process.

Table 3: Paint Defects

No	Defects	Total Qty	Reject PPM	Cumulative
1	Runs	0	0	48.26
2	Thin paint	0	0	78.98
3	Dirt	680	38	93.61
4	Craters	640	80	96.38
5	Scratches	665	52	100.00

The analysis shows that 'Runs' is the leading defect type, responsible for 48.26% of total rejections, with a reject rate of 23,125 PPM as indicated in Figure 3. 'Thin paint' follows, contributing 31% to overall defects and a reject of 14,722 PPM. Combined, these two issues account for nearly 79% of paintshop rejections, highlighting them as the main drivers of rework and material waste in the paint process. Defects such as 'runs', 'thin paint', and 'craters' are caused by irregular process parameters, such as the temperature, pressure, and paint viscosity.

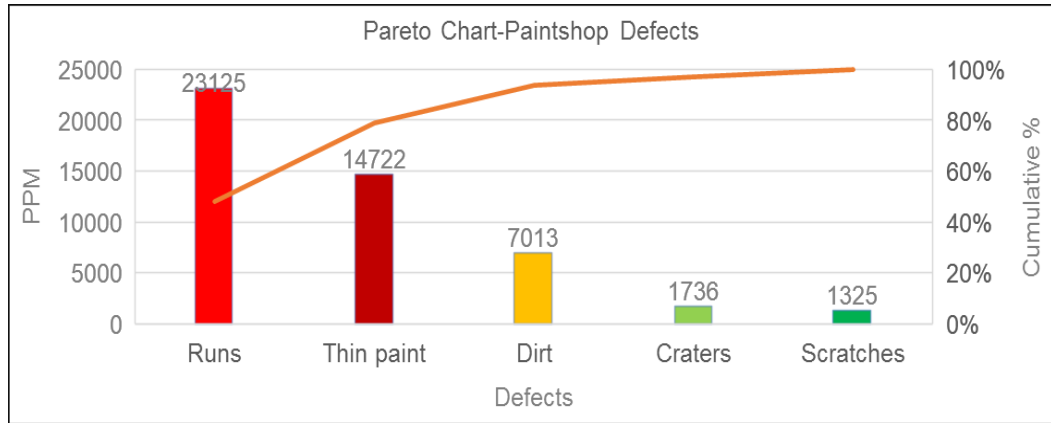


Figure 3: The Pareto Chart of the Paintshop

To determine the optimal conditions that would yield maximum painting productivity, minimize rework and scrap while optimizing FTT%, an ML model is developed. A total of 92 data entries consisting of hourly data of the temperature, pressure, viscosity, and humidity in the paintshop booth are monitored and collected for model development.

3.1 Random Forest Regression Model Result.

Since the data is measured hourly and aggregated into daily data by calculating the daily mean, mode, standard deviation, and variance. The evaluated data are the features of the model, and are 12 features in total, while the model output is good products (FFT), the number of rework and scraps. Feature importance is performed on the normalized features using SHapley Additive exPlanations (SHAP) to rank the features in order of importance as represented in Figure 4. It ranks each feature based on how it affects each final prediction, the relative importance of each feature, and analyses the model's dependence on interactions among them [49, 50].

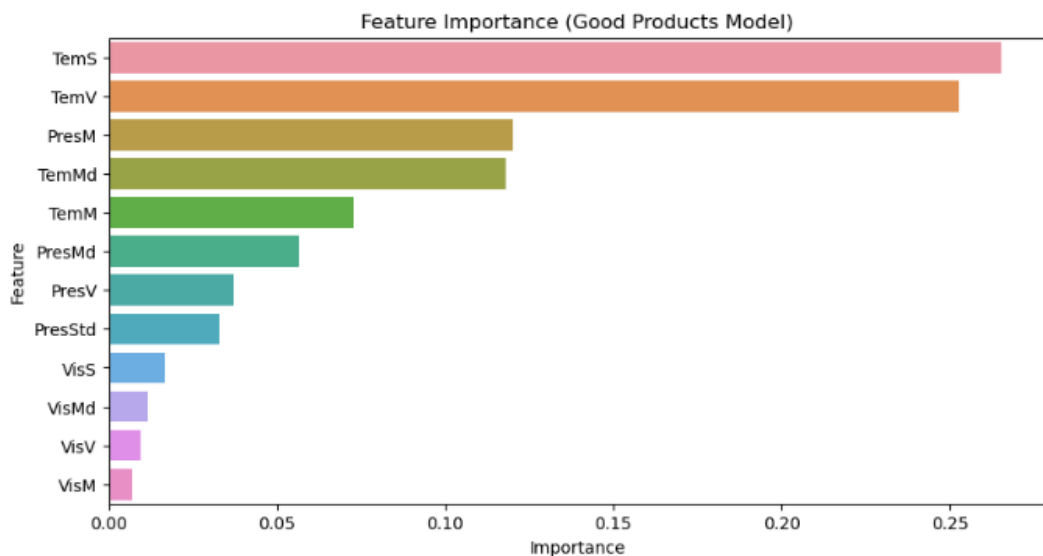


Figure 4: Feature Importance (Good Products Model)

The random forest regression model developed was trained, and the root mean square error (RMSE) in predicting good product, rework, and scraps is 0.2914439, 0.296685, and 0.43, respectively. The optimisation of the objective function that maximizes the good products and minimizes rework and scrap with values 1.22983826, 1.5478891, 6.06431159 for the variance of Temperature, standard deviation of the temperature, and mean pressure as illustrated in Figure 5.

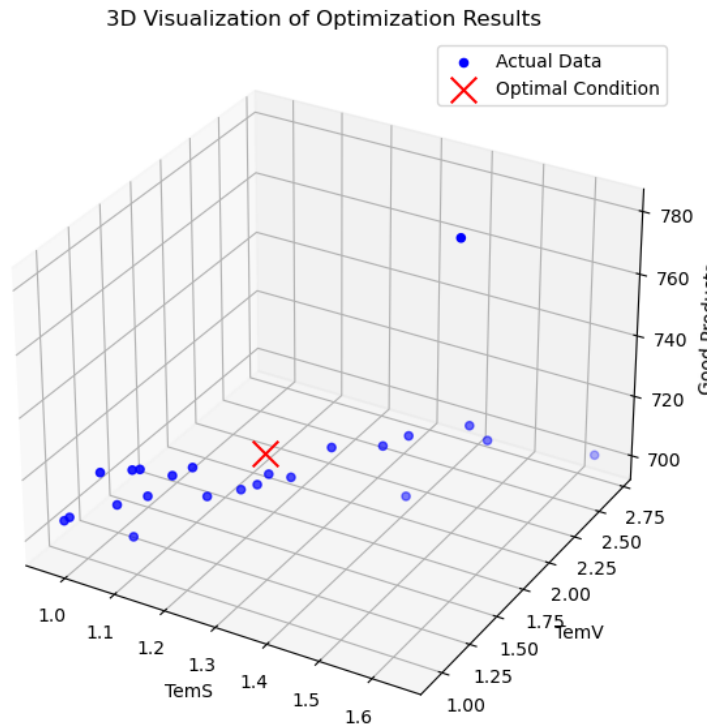


Figure 5: Optimization Result Chart

When the 12 features are used for modelling and optimization, the optimal values of the 12 features are illustrated in Table 4.

Table 4: Optimal Values for 12 Parameters using Random Forest Algorithm

TemM	VisM	PressM	TemS	TemMd	TemV
23.5664	1.2298	23.1304	1.5479	18.0399	0.7231
VisS	VisMd	VisV	PresStd	PreMd	PresV
18	0.5451	6.0641	0.3909	6.1522	0.2341

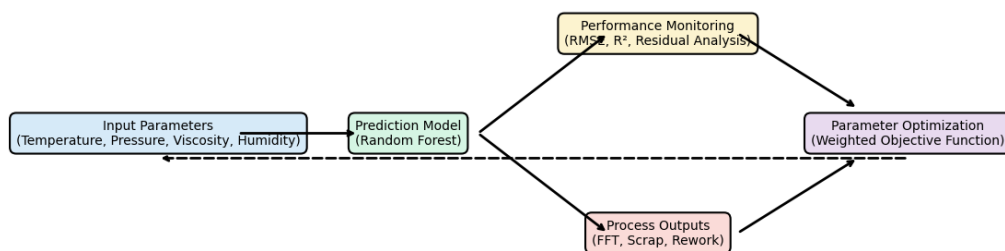


Figure 6: Closed-Loop Model Validation and Optimization Framework

A closed-loop validation framework is implemented where model predictions are continuously evaluated using performance metrics, as shown in Figure 6. Based on these results, process parameters are iteratively adjusted through optimization, forming a feedback loop that improves process performance over time.

To further validate the result of the model (the optimal values of the features), another deep learning algorithm, a reinforcement learning model with Q-learning parameters, is developed. The Q-learning parameters that determine the actions ('increase', 'decrease', or 'stay') are the learning rate, α , which is 0.1, the discount factor, γ , which is 0.9, and the exploration rate, ϵ , which is 0.2. while the episode of learning is 5000, as illustrated in Figure 7. The model discretized the features, and the optimal values for the parameters using a random forest regressor with reinforcement learning are presented in Table 5.

Table 5: Optimal Values for the Parameters Using Q-Reinforcement Learning

TemM	VisM	PressM	TemS	TemMd	TemV
23.94	1.04	22.90	1.11	18.19	0.89
VisS	VisMd	VisV	PresStd	PreMd	PresV
18	0.68	6.15	0.09	6.0	0.08

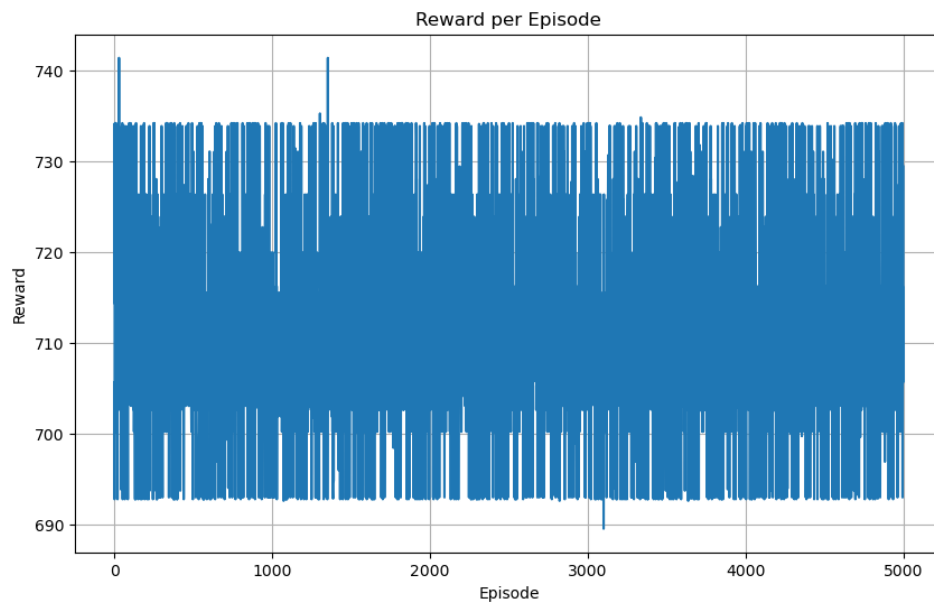


Figure 7: The Model's Episode

The RMSE for the Q-learning model is 0.05439, indicating that the model's output is more accurate in learning the optimal level that could yield the maximum good painted part. These parameter levels are noted and fed into the system for optimal performance, while other external factors are examined using a Cause-and-Effect diagram.

3.0 DISCUSSION

The results of the study indicate that the operating parameters essentially contribute to the number of good products, scrap, and rework achieved during production. Pareto diagram of the defects shows that three of the top 5 defects causing rework and scraps were a result of painting operating parameters, such as the temperature, pressure, and viscosity.

Determining the optimal threshold for each parameter helps to drastically reduce the number of reworks and scraps by essentially compelling painting operations to the optimal values. The 2 different models were developed and trained using the production data and the hourly data to understand and determine the parameter threshold that optimizes the painting of the good part. The random forest regressor model and Q-reinforcement learning were developed. The RMSE for each model indicates the model's degree of accuracy, and the Q-reinforcement learning algorithm shows a more accurate result based on the RMSE value. The Random Forest model demonstrated strong predictive performance, achieving an R^2 value of approximately 0.72 for FFT prediction, indicating that over 70% of the variance in the target variable is explained by the model. The corresponding RMSE values were relatively low and consistent between cross-validation and test datasets, suggesting good generalization and minimal overfitting.

To evaluate the effectiveness of the machine learning model, its performance was compared with baseline approaches, including a mean predictor and a linear regression model, as illustrated in Table 6. The mean predictor, which assumes no relationship between input variables and output, resulted in high prediction error and near-zero R^2 , confirming its limited usefulness.

Table 6: Performance Evaluation against some Baseline

Model	RMSE	R^2
Mean	High	~0
Linear	Medium	~0.3–0.5
Random Forest	Lowest	~0.7+

The linear regression model showed moderate improvement, capturing some linear relationships between process parameters and outputs. However, its performance was significantly lower than that of the Random Forest model, indicating that the underlying relationships are likely nonlinear and involve interactions between variables.

The Random Forest model outperformed both baseline methods, achieving the lowest RMSE and highest R^2 values. This demonstrates its ability to effectively model complex, nonlinear dependencies in the data, making it more suitable for process optimization tasks. The superior performance of the Random Forest model suggests that process variables such as temperature, pressure, viscosity, and humidity interact in a nonlinear manner to influence FFT, scrap, and rework. Traditional linear models are unable to fully capture these interactions, whereas ensemble learning methods provide improved predictive capability, enabling more reliable optimization of process parameters.

The parameter's optimal values are used as benchmarks in adjusting the system parameters for improved performance. The model's parameters were validated by adjusting the system parameters to check the accuracy of the model over a month period while implementing some system improvements to enhance production performance. In June, the target was to paint 14,400 units, but only 13,834 units were painted, achieving a First Time Through (FTT) rate of 95.03%. 692 units needed rework, and 24 units were scrapped. In contrast, August saw a total of 16,395 units painted out of a target of 16,560, which shows a much greater FTT rate of 99.00% as indicated in Table 7. There was a sharp decrease in reworks, which fell to 160 units, while scrapped items were limited to only 5. This improvement suggests that the implemented solutions positively influenced production performance. The data further indicate improved stability and efficiency in the production process, resulting in a lower risk of defects and mistakes.

Table 7: Production Summary Report for August 2024

Total Monthly Target	Monthly Output	Monthly Rework	Monthly Scrap	Monthly FFT%	Monthly Rework%	Monthly Scrap%
16560	16395	160	5	99.00	0.97	0.03

Clifton and Laber [51] describe Q-learning as an algorithm originally developed to incrementally estimate the best strategy in problems involving decisions over an infinite time horizon. This is a robust algorithm capable of determining the optimal value that optimizes a condition or actions. Q-learning is a discrete reinforcement learning algorithm that selects actions based on a strategy informed by rewards and interactions with the environment, aiming to estimate the optimal Q-function [52]. A further study into the possible causes that could disrupt the painting parameters during operations is conducted using the Cause-and-Effect diagram, as indicated in Figure 8.

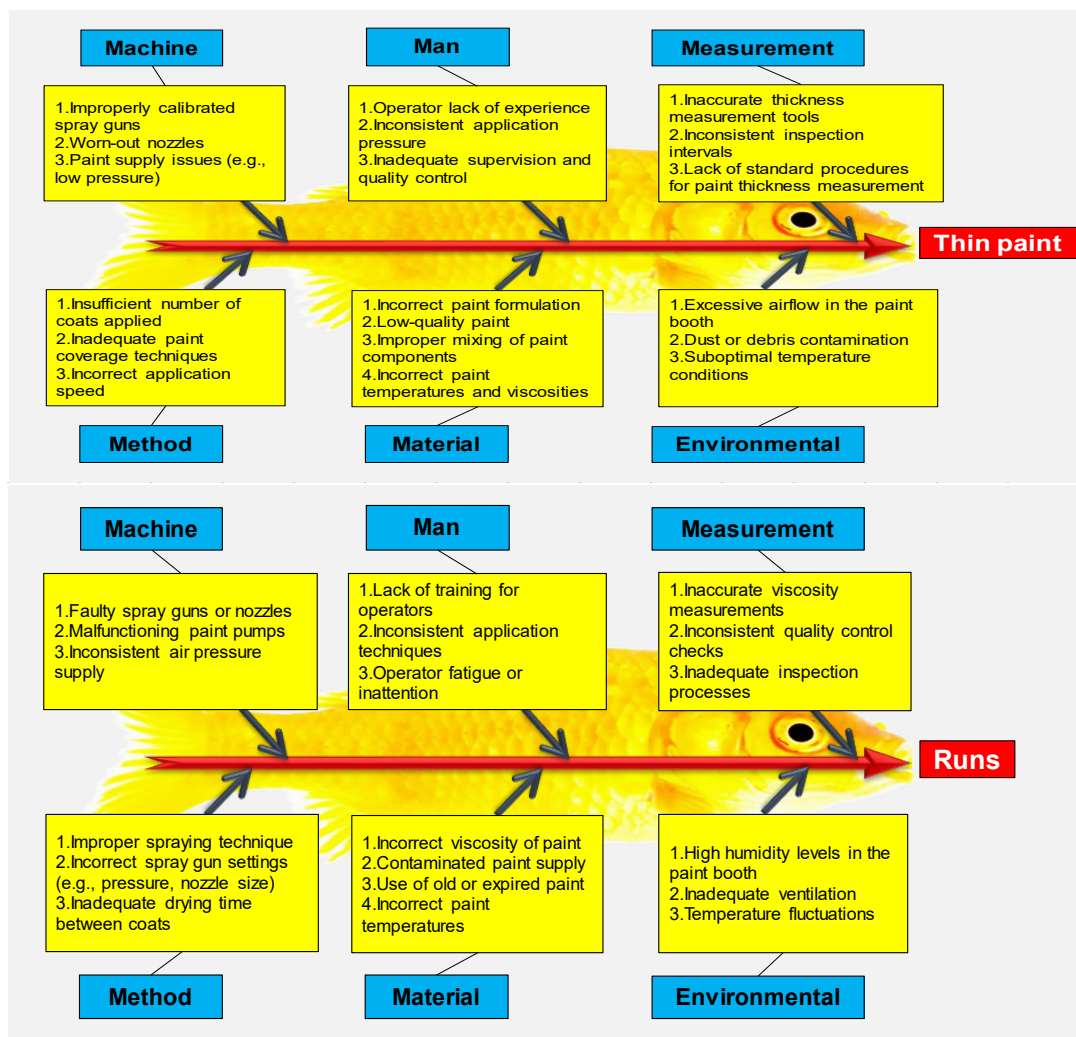


Figure 8: The Cause-and-Effect Diagram

Several potential causes of the defects were identified using the fishbone analysis to determine other causes of the defects, with the influence of the painting parameters to ensure that other factors are also addressed. These causes were examined under factors

such as Machine, Manpower, Measurement, Method, Material, and Environment to ensure all these factors are addressed, and the significance of the optimal threshold determined by the model is achieved. Environmental factors within the paint booth—such as temperature, humidity, and downdraft airflow—significantly influenced the quality of paint application. Variations in these conditions contributed to defects such as runs and insufficient paint coverage.

4.0 CONCLUSION

The results of this study demonstrate that the application of a data-driven machine learning approach can significantly improve production performance. By leveraging the optimized process parameters derived from both the Random Forest and Q-learning models, notable improvements were observed across key performance indicators.

From a quantitative perspective, the First Time Through (FTT) rate improved from 95.03% in June to 99.00% in August, representing a substantial increase in process efficiency. This improvement was accompanied by a significant reduction in defects: rework decreased from 692 units to 160 units, while scrap was reduced from 24 units to 5 units. These results clearly indicate that the implementation of optimized parameters, while addressing the underlying causes, contributed to enhanced process stability and product quality.

From an industrial standpoint, these improvements have important implications. A higher FTT rate directly translates to increased productivity, reduced operational costs, and improved resource utilization. The reduction in rework and scrap minimizes material waste, labor costs, and downtime, thereby enhancing overall manufacturing efficiency. Furthermore, the developed framework demonstrates the potential of machine learning models to support data-driven decision-making and process optimization in real-world production environments.

However, several limitations should be acknowledged. The dataset used for model development was relatively small and covered a limited time, which may affect the robustness and generalizability of the model. Additionally, the model was built using a specific set of process variables (e.g., temperature, pressure, humidity, and viscosity), and other potentially influential factors such as material variability, operator effects, or equipment conditions were not included. As a result, the model may not fully capture all sources of process variation.

Future research should focus on expanding the dataset over a longer time horizon and incorporating additional process variables to improve model accuracy and robustness. Furthermore, integrating real-time data acquisition and deployment within a live production environment would enable continuous monitoring and adaptive optimization, aligning with Industry 4.0 principles.

In conclusion, the study highlights the effectiveness of machine learning-based optimization in improving manufacturing performance, while also providing a foundation for further development toward more scalable and intelligent production systems.

ACKNOWLEDGEMENT

The University of South Africa (UNISA), through its Department of Industrial Engineering, provided full support for this study.

REFERENCES

- [1] Akafuah, N.K., et al., Evolution of the automotive body coating process—A review. *Coatings*, 2016. 6(2): p. 24.
- [2] Eley, R.R., Applied rheology and architectural coating performance. *Journal of Coatings Technology and Research*, 2019. 16(2): p. 263-305.
- [3] Brega, D., O. Trifonov, and I. Voronko, Protective coatings. 2019.
- [4] Maile, F.J., Colorants in coatings. *Physical Sciences Reviews*, 2021. 6(11): p. 707-789.
- [5] Konstantinidis, F.K., et al., A technology maturity assessment framework for Industry 5.0 machine vision systems based on systematic literature review in automotive manufacturing. *International Journal of Production Research*, 2023: p. 1-37.
- [6] Okokpujie, I.P., et al., Effect of coatings on mechanical, corrosion, and tribological properties of industrial materials: a comprehensive review. *Journal of Bio-and Tribo-Corrosion*, 2024. 10(1): p. 2.
- [7] Punetha, M., et al., Industrial Scale Production, Commercialization, and Global Market of Functionalized Carbon Nanostructures, in *Handbook of Functionalized Carbon Nanostructures: From Synthesis Methods to Applications*. 2024, Springer. p. 1-58.
- [8] Bonaduce, I., et al., Conservation issues of modern oil paintings: a molecular model on paint curing. *Accounts of chemical research*, 2019. 52(12): p. 3397-3406.
- [9] Hakan Behic, A., Business development approaches for the painting subcontracting using innovations in automation technologies. 2021.
- [10] Gabbar, H.A. and M. Idrees, ARSIP: Automated robotic system for industrial painting. *Technologies*, 2024. 12(2): p. 27.
- [11] Yadav, D. and J. Dutta, Advancing environmental sustainability: recent trends and developments in treatment methods for paint industry wastewater. *Journal of Water Process Engineering*, 2024. 61: p. 105290.
- [12] Viktorová, N., A. Szarka, and S. Hrouzková, Recent developments and emerging trends in paint industry wastewater treatment methods. *Applied Sciences*, 2022. 12(20): p. 10678.
- [13] Jhamb, S., et al., A review of computer-aided design of paints and coatings. *Current Opinion in Chemical Engineering*, 2020. 27: p. 107-120.
- [14] Pendar, M.-R., et al., Review of coating and curing processes: Evaluation in the automotive industry. *Physics of Fluids*, 2022. 34(10).
- [15] Chen, X., et al., Development and application of intelligent coating technology: a review. *Coatings*, 2024. 14(5): p. 597.
- [16] Sahoo, B.N. and B. Kandasubramanian, Recent progress in fabrication and characterisation of hierarchical biomimetic superhydrophobic structures. *Rsc Advances*, 2014. 4(42): p. 22053-22093.
- [17] Liu, Q. and M.W. Urban, Stimulus-responsive macromolecules in polymeric coatings. *Polymer Reviews*, 2023. 63(2): p. 289-323.
- [18] Buratto, W.G., et al., A review of automation and sensors: parameter control of thermal treatments for electrical power generation. *Sensors*, 2024. 24(3): p. 967.
- [19] Pendar, M.-R., et al., Enhancing Automotive Paint Curing Process Efficiency: Integration of Computational Fluid Dynamics and Variational Auto-Encoder Techniques. *Sustainability*, 2025. 17(7): p. 3091.
- [20] Corbu, E.C., A.-M. Nitescu, and E. Edelhauser, Optimizing Correction Factors on Color Differences for Automotive Painting Services. *Sensors*, 2024. 24(24): p. 8213.
- [21] Guettler, N., et al. A self-programming painting cell. SelfPaint: Simulation-based path generation with automated quality control for painting in small lot sizes. in *Advances in Automotive Production Technology—Theory and Application: Stuttgart Conference on Automotive Production (SCAP2020)*. 2021. Springer.
- [22] Malik, P. and D. Sarkhel, Advancements in bio-based materials and low-VOC formulations: paving the way for sustainable innovation in the coatings industry. *Journal of coating technology and innovation*, 2024(2): p. 2.
- [23] Zhang, R., Environment-aware production scheduling for paint shops in automobile manufacturing: A multi-objective optimization approach. *International journal of environmental research and public health*, 2018. 15(1): p. 32.
- [24] Wang, J., J. Li, and N. Huang. Optimal scheduling to achieve energy reduction in automotive paint shops. In the *International Manufacturing Science and Engineering Conference*. 2009.
- [25] Howells, E.B., Measuring temperature. *Anaesthesia & Intensive Care Medicine*, 2018. 19(10): p. 563-567.
- [26] Cavalcante, E.S., et al., Automotive painting process: Minimizing energy consumption by using adjusted convective heat transfer coefficients. *Progress in Organic Coatings*, 2020. 140: p. 105479.
- [27] Honzicek, J., Curing of air-drying paints: a critical review. *Industrial & engineering chemistry research*, 2019. 58(28): p. 12485-12505.
- [28] Müller, B. and M. Schackmann, Coatings formulation. 2023: Lack in Vincentz GmbH & Company KG.
- [29] Anderson, A.M., B.A. Bruno, and L.S. Smith, Viscosity measurement. *AA*, 2014. 2(t3): p. 809.

- [30] Leblanc, G., R. Secco, and M. Kostic, Viscosity measurement, in *Mechanical Variables Measurement-Solid, Fluid, and Thermal*. 2023, CRC Press. p. 11-1-11-24.
- [31] Wypych, G., *Handbook of Rheological Additives*. 2022: Elsevier.
- [32] Wen, Y., D.J. Jaeger, and P.R. Pagilla, Uniform coverage tool path generation for robotic surface finishing of curved surfaces. *IEEE Robotics and Automation Letters*, 2022. 7(2): p. 4931-4938.
- [33] Rajagopal, K., Remarks on the notion of "pressure". *International journal of non-linear mechanics*, 2015. 71: p. 165-172.
- [34] Poozesh, S., N. Akafuah, and K. Saito, Effects of automotive paint spray technology on the paint transfer efficiency—a review. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 2018. 232(2): p. 282-301.
- [35] Hamburg, H.R., *Hess's Paint Film Defects: Their Causes and Cure*. 2012: Springer Science & Business Media.
- [36] He, W., et al., Review of flow control for atomization and sprays: Application section. *International Journal of Heat and Mass Transfer*, 2025. 236: p. 126300.
- [37] Chan, T., J. D'arcy, and R. Schreck, High-solids paint overspray aerosols in a spray-painting booth: particle size analysis and scrubber efficiency. *American Industrial Hygiene Association Journal*, 1986. 47(7): p. 411-417.
- [38] Boulos, M.I., et al., Plasma Spray Process Integration, in *Handbook of Thermal Plasmas*. 2023, Springer. p. 1313-1422.
- [39] Diwekar, U.M., *Introduction to applied optimization*. Vol. 22. 2020: Springer Nature.
- [40] Zunaidi, I., Parameter settings for an automated gantry-robot painting system using a 3-gun atomization spray method for an anti-static coating process. *American Journal of Engineering and Applied Sciences*, 2019. 12(4).
- [41] Pendar, M.-R., S. Cândido, and J.C. Páscoa, Optimization of painting efficiency applying unique techniques of high-voltage conductors and nitrotherm spray: Developing deep learning models using computational fluid dynamics dataset. *Physics of Fluids*, 2023. 35(7).
- [42] Pendar, M.-R. et al., Enhancing automotive paint curing process efficiency: Integration of computational fluid dynamics and variational Auto-Encoder techniques. *Sustainability*, 2025. 17(7): p. 3091.
- [43] Kwofie, F., et al., Transmission infrared microscopy and machine learning applied to the forensic examination of original automotive paint. *Applied Spectroscopy*, 2022. 76(1): p. 118-131.
- [44] Torrent Solà, E., *Process standardization for car interior paint production: a comprehensive study on the paint production transfer from Offenbach to Barcelona*. 2024.
- [45] Ao, Y., et al., The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. *Journal of Petroleum Science and Engineering*, 2019. 174: p. 776-789.
- [46] Kim, S.H. and F. Boukouvala, Machine learning-based surrogate modeling for data-driven optimization: a comparison of subset selection for regression techniques. *Optimization Letters*, 2020. 14(4): p. 989-1010.
- [47] Arulkumaran, K., et al., A brief survey of deep reinforcement learning. *arXiv preprint arXiv:1708.05866*, 2017.
- [48] Arulkumaran, K., et al., Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 2017. 34(6): p. 26-38.
- [49] Meng, Y., et al., What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, 2020. 16(3): p. 466-490.
- [50] Lee, Y.-G., et al., Shap value-based feature importance analysis for short-term load forecasting. *Journal of Electrical Engineering & Technology*, 2023. 18(1): p. 579-588.
- [51] Clifton, J. and E. Laber, Q-learning: Theory and applications. *Annual Review of Statistics and Its Application*, 2020. 7(1): p. 279-301.
- [52] Tan, F., P. Yan, and X. Guan. Deep reinforcement learning: From Q-learning to deep Q-learning. In *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part IV* 24. 2017. Springer