Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

AI-BASED ASSEMBLY LINE OPTIMIZATION IN AERONAUTICS: A SURROGATE AND GENETIC ALGORITHM APPROACH

Maryam Saadi[1,2], Vincent Bernier[1], Gregory Zacharewicz[2], and Nicolas Daclin[2]

[1]Digital Design, Manufacturing and Services, Airbus Helicopters, Marignane, FRANCE
[2]SyCoIA, Institut Mines Telecom, Ales, FRANCE

ABSTRACT

Industrial configuration planning requires testing many setups, which is time-consuming when each scenario must be evaluated through detailed simulation. To accelerate this process, we train a multi-layer perceptron to predict key performance indicators quickly, using it as a surrogate model. However, classical regression metrics such as mean squared error, mean absolute error, and root mean squared error do not reflect prediction quality in all situations. To solve this issue, we introduce a classification-based evaluation strategy. We define acceptable prediction margins based on business constraints, then convert the regression output into discrete classes. We assess model performance using precision and recall. This approach reveals where the model makes critical errors and helps decision-makers at airbus helicopters trust the artificial intelligence predictions.

1 INTRODUCTION

In the aerospace industry, production systems face increasing pressure to adapt to shifting customer demands and operational constraints. In helicopter assembly, planners must evaluate numerous configuration scenarios to ensure timely delivery and cost control. This evaluation traditionally relies on detailed simulations that replicate real-world operations such as component preparation, workforce planning, and final assembly (Forrester et al. 2008).

However, these simulations are often too slow to support time-sensitive decisions. At Airbus Helicopters, testing thousands of configurations may require several days (Saves et al. 2024). This delay hinders responsiveness and limits the ability to explore alternative solutions rapidly.

To accelerate decision-making, we propose replacing the simulation step with an AI-based surrogate model. These models, including neural networks, can approximate simulation outcomes with much faster computation (Saadi et al. 2024). These predictions are fast but they must remain accurate and reliable—especially near thresholds that guide operational decisions.

We designed a hybrid framework that combines a Multi-Layer Perceptron (MLP) and a Genetic Algorithm (GA). The MLP is trained on synthetic data generated from past simulation runs (Saadi et al. 2024). The GA explores the parameter space to identify high-performing configurations based on the surrogate predictions.

The challenge here is to evaluate the trustworthiness of these predictions. Standard regression metrics such as Mean Absolute Error (MAE) and Root Mean Absolute Error (RMSE) provide general accuracy scores but often hide local errors near critical boundaries (Suarez et al. 2024). By critical local errors, we refer to inaccuracies near operational thresholds, such as a 90% satisfaction rate, where even small deviations may lead to poor decisions.

The dataset used in this context is complex and high-dimensional. It is difficult to identify which types of customer requests lead to inaccurate predictions. These hidden errors can result in poor decisions and impact operational performance.

Figure 1 illustrates the industrial context of our study, showing the structure of the MECA 4.0 workshop and the two production flows handled. MECA refers to Mechanics, and the term 4.0 highlights a digitalized and reconfigurable workshop architecture that combines serial production and the maintenance, repair, and overhaul (MRO) flows. In addition, traditional optimization methods such as response surface methodology (Montgomery 2017) have been widely applied in industrial design of experiments, but they are often limited by the number of input dimensions and assume smooth functional responses. This paper makes two contributions:

- It introduces a surrogate-assisted GA framework that enables fast and effective optimization of workshop configurations.
- It proposes a classification-based evaluation method that focuses on business-relevant decision thresholds and improves interpretability.

Unlike traditional regression evaluations, our method targets high-risk regions in the decision space. Applied at Airbus Helicopters, it reduces decision time and increases confidence in AI-supported planning.

Our research extends previous studies on simulation with machine learning, surrogate-based optimization, and evolutionary algorithms. We integrate these methods into one framework for real-time optimization in manufacturing. We also introduce a classification-based validation strategy. This strategy addresses the problem of model trust. It provides a step toward surrogate systems that experts can apply with confidence in industrial decision-making.

2 STATE OF THE ART

2.1 Machine Learning and Simulation in Manufacturing

The integration of machine learning (ML) with simulation methodologies has become increasingly relevant for optimizing decision-making in complex manufacturing environments. As noted by (Hürkamp et al. 2021), two fundamental approaches exist for combining these technologies. The first approach leverages simulation to support ML—typically by generating synthetic training data or enriching AI models with simulation-derived features. The second approach uses ML to support or replace simulations, with the goal of reducing computational cost and enabling faster evaluations.

In this context, several studies have investigated the use of neural networks as surrogate models that emulate simulation outputs. For instance, Jin (2011) advocated for ML-driven "metamodels", while Pestourie et al. (2020) proposed an active learning framework that dynamically selects simulation samples to refine a surrogate in a smart factory setting. Other contributions have emphasized building comprehensive surrogate datasets in advance, which can then be used directly in optimization routines without iterative retraining.

2.2 Surrogate Models for Optimization

Surrogate-based optimization refers to strategies where simplified models are substituted for computationally expensive evaluation functions to expedite optimization tasks. This concept has a long-standing history in engineering, particularly in fields such as aerodynamic design and process optimization. Surrogates like Kriging (Gaussian process regression), radial basis functions, and polynomial approximations have traditionally been employed for this purpose (Forrester et al. 2008). More recently, deep learning models such as neural networks have gained traction due to their scalability and ability to model high-dimensional spaces. Pereira et al. (2025), among others, highlighted the advantages of deep learning-based surrogates in fine-tuning complex process parameters, noting their ability to achieve faster convergence to optimal configurations than traditional simulation-driven approaches. These surrogate models also allow for greater computational efficiency, enabling heuristic optimization techniques (such as GA and particle swarm optimization) to explore larger solution spaces within fixed computational budgets. A representative example is the SMT 2.0 toolkit developed by Saves et al. (2024), which introduces advanced Gaussian Process models tailored for surrogate-based optimization. Despite these advances, most frameworks emphasize modeling accuracy and flexibility for complex variable types, but they remain computationally demanding and less adapted to real-time industrial deployment. This limitation reduces their applicability in contexts where decisions must be made rapidly and under operational constraints.

2.3 Evolutionary Algorithms and Surrogate Assistance

Evolutionary algorithms, particularly GAs, are widely used to solve complex optimization problems in manufacturing, such as job shop scheduling and facility layout planning. However, their practical application is often limited by the computational cost of evaluating each solution, especially when simulations are used for fitness evaluations. To address this bottleneck, researchers have explored surrogate-assisted evolutionary optimization, wherein a learned model approximates the simulation outcomes to accelerate the evaluation process.

For example, a study by Stander et al. (2022) demonstrated the benefits of incorporating a neural network into a GA for optimizing a production system. The surrogate-assisted algorithm outperformed the standard GA in terms of scalability and efficiency, enabling broader exploration of the solution space. Similarly, prior work by Li et al. (2022) and Paape et al. (2024) reported improvements in energy optimization and manufacturing cell design by combining GAs with accurate surrogate models.

Despite these advances, most studies remain focused on academic test cases or simplified scenarios. Their effectiveness in industrial environments is less documented, particularly when decision thresholds or operational risks must be taken into account. Moreover, surrogate models used in these approaches can mislead the optimizer if local prediction errors occur, especially near critical regions of the search space. This risk highlights the need for further research on validation strategies and robust integration of surrogates within evolutionary algorithms.

2.4 Trust and Risk in Surrogate Models

Beyond accuracy, the trustworthiness of surrogate models is a vital concern—especially in operational contexts where incorrect predictions can lead to costly or unsafe outcomes. High overall accuracy metrics (e.g., R^2 or mean squared error) may not be sufficient if the model fails in regions critical to decision-making.

To address this, we apply a classification-based error analysis framework. Instead of evaluating the model solely on continuous predictions, we categorize outputs as either acceptable or unacceptable based on predefined thresholds and assess performance using precision and recall. High precision ensures that most configurations predicted as successful are indeed valid, while high recall indicates that the model captures the majority of viable options. This approach aligns with the goal of building trustworthy AI systems for industrial use, as seen in aerospace applications (Ducoffe et al. 2024), where reliability near safety margins is paramount.

This evaluation method also complements uncertainty quantification and Bayesian optimization by providing an interpretable assessment of model reliability. By explicitly framing the surrogate as a decision-support tool rather than a black-box predictor, our methodology enhances transparency and facilitates confident adoption in industrial workflows.

3 METHODOLOGY

3.1 Industrial Workflow: The MECA 4.0 Workshop

The MECA 4.0 workshop at Airbus Helicopters is dedicated to the assembly of key helicopter components, including the Main Gearbox (MGB) and the Main Rotor Head (MRH). As shown in Figure 1, the green and orange flows represent serial and MRO production, respectively. Parts are first prepared in the ZPP area, then pass through sub-assembly zones (PSE, CA) before reaching final assembly or spare stock. The flow is regulated by CONWIP buffers across the line. A complete list of abbreviations is available in Appendix A.

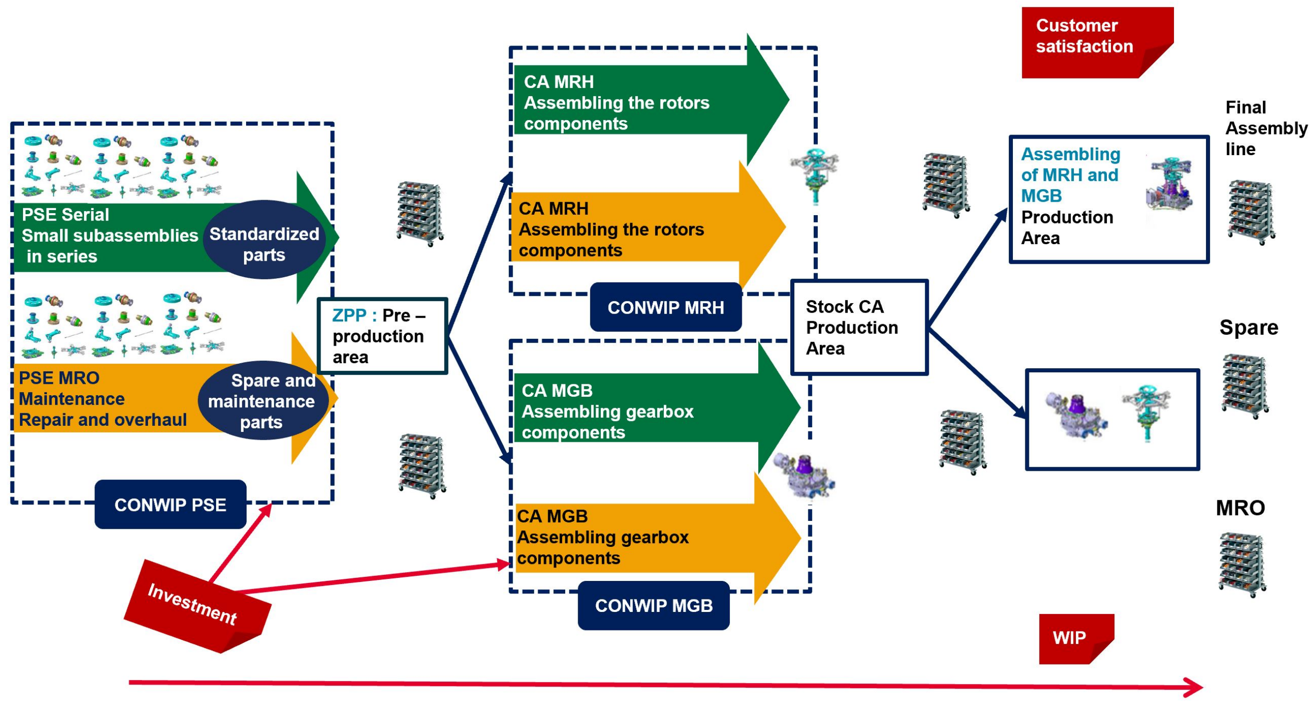

Figure 1: Model of the MECA 4.0 workshop.

3.2 Data Generation with Simulation

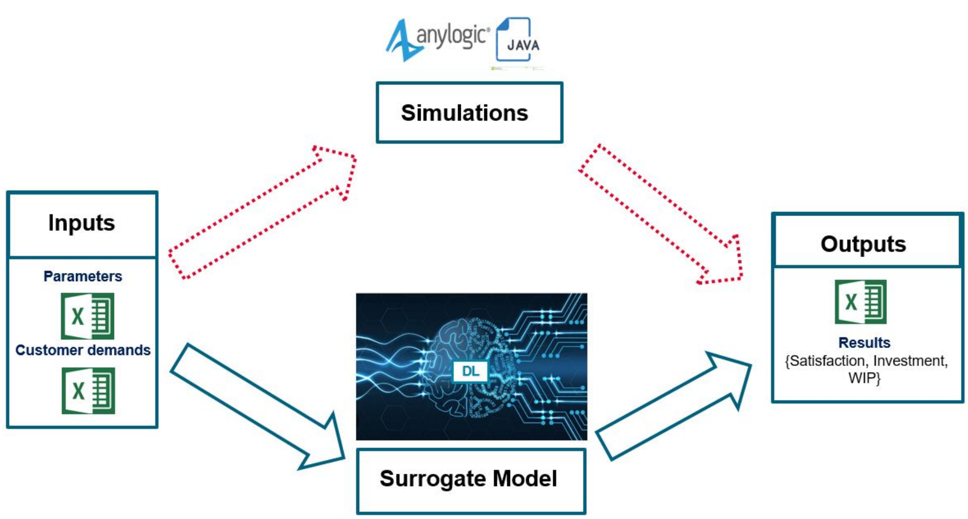

Figure 2: Surrogate model substituting simulation for fast KPI estimation.

To simulate the workshop behavior, a hybrid model was developed using AnyLogic (version 8.5.1 Professional) as shown in figure 2. It combines discrete event and agent-based modeling to capture complex production logic. Development of the model required between six months and two years. Once finalized, a Java-exported version allowed parameter adjustments without reopening the graphical interface.

We use the simulation model to calculate the key performance indicators (KPIs): WIP, investment, and customer satisfaction. The evaluation of 20 different configurations requires approximately four days. This computational cost limits its use in fast decision-making.

3.3 Surrogate Model Inputs and Outputs

To reduce evaluation time, we developed a surrogate model based on a Multi-Layer Perceptron (MLP). This model was designed to substitute the original simulation model. It uses the same inputs and produces the same outputs as the AnyLogic simulation. The MLP receives two types of input:

- **Customer demands:** helicopter type, delivery deadlines, and expected production volumes.
- **Workshop parameters:** structure and capacity, number of workers, production stations, and available resources.

The model predicts the following three KPIs, which match those computed by the simulation:

- **Customer satisfaction:** normalized ratio of on-time deliveries.
- **Work-in-Progress (WIP):** total value of unfinished assemblies on the shop floor, representing capital tied up during production.
- **Investment:** total cost required to implement the selected configuration.

These predicted KPIs are used to train the MLP and to evaluate candidate configurations during the optimization phase.

3.4 Optimization with Genetic Algorithm

A genetic algorithm was integrated with the surrogate model to identify optimal workshop configurations. The GA is based on evolutionary principles such as selection, crossover, and mutation.

Each iteration begins with a population of configurations generated by combining parameter files with client request files. These configurations are evaluated using the surrogate model, which estimates the KPIs. A fitness score is assigned to each candidate solution. The best-performing configurations are selected for crossover and mutation to create the next generation.

This process runs for 200 iterations with up to 30 solutions per generation, producing around 6000 candidate configurations. All results are stored in a database as triplets:{customer request, configuration parameters, predicted KPIs}.

4 ANALYSIS OF PREDICTION ERRORS AND MODEL PERFORMANCE

An MLP was developed as a fully connected feedforward neural network. It processes customer requests and configuration parameters using an input layer followed by four dense layers, starting with 256 neurons and progressively reducing to 128, 64, and 32 neurons. This structure allows the network to capture complex relationships and refine features. To mitigate overfitting, a dropout rate of 25% is applied in selected layers. This architecture was selected after empirical testing to balance model complexity and training stability, ensuring the model captured non-linear interactions without overfitting.

To assess prediction accuracy, we used two regression metrics: MAE and RMSE. MAE provides a direct interpretation of the average prediction error in the same units as the target variable, which makes it intuitive and robust to outliers. RMSE, gives more weight to larger deviations, making it particularly useful for highlighting significant mispredictions in critical scenarios.

4.1 MLP Prediction Analysis

To evaluate the model, we examine prediction errors for each KPI using MAE and RMSE. Table 1 shows that the model predicts Investment with the highest accuracy, with both MAE and RMSE below 0.4%. This is expected, as the relationship between inputs and Investment is mostly linear and deterministic, making it easier for the model to learn.

WIP predictions are less accurate but still acceptable. The model shows moderate dispersion, especially for high WIP values. This reflects the complexity of WIP, which depends on multiple interactions in the shop floor such as bottlenecks and queue lengths.

Customer Satisfaction is the most challenging to predict. This variability suggests that customer satisfaction depends on non-linear effects and interactions that are harder to model directly, such as cascading delays and configuration mismatches.

Table 1: Error metrics for wip, investment, and customer satisfaction.

Metric	WIP	Investment	Customer Satisfaction
MAE	3.51%	0.23%	4.99%
RMSE	4.84%	0.35%	7.14%

Table 2: Prediction accuracy within tolerance margins for each KPI.

KPI	≤5% Error (%)	≤10% Error (%)
WIP	75.33%	95.44%
Investment	99.91%	100.00%
Customer Satisfaction	53.19%	83.11%

Table 2 provides a more detailed view of prediction quality. It shows the percentage of predictions that fall within a 5% and 10% error margin.

The model performs well on Investment, with nearly all predictions under a 5% error. For WIP, 75% of the predictions are within 5% error, and over 95% are under 10%. These results confirm that the model captures the main patterns despite WIP's complexity.

Customer Satisfaction is more difficult. Only 53% of predictions fall within the 5% margin, and about 83% within the 10% range. This confirms the higher variability observed earlier and highlights the need for deeper analysis, especially on configurations with large satisfaction errors.

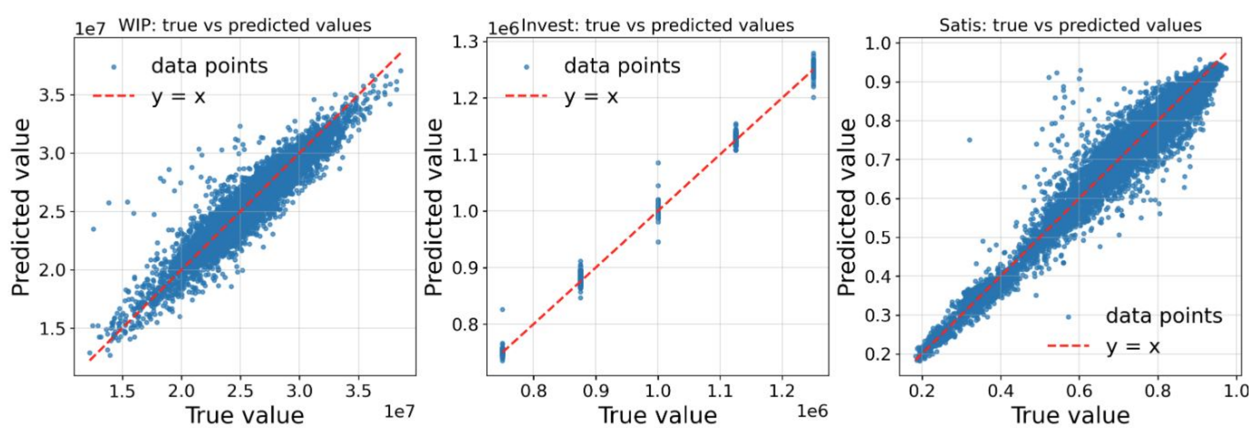

Figure 3: Evaluation of model predictions: true vs. predicted values for wip, investment, and customer Satisfaction.

Figures 3 helps visualize these trends. The scatter plots show how predictions align with actual values. Investment values lie very close to the diagonal (ideal fit), while WIP shows more spread. Satisfaction predictions clearly deviate more, especially for extreme values, revealing that the model struggles with underfitting or overfitting in some ranges.

This analysis shows that global metrics are not sufficient to assess model reliability. Airbus has defined a stricter requirement: for Customer Satisfaction, at least 90% of the predictions must have less than 5% error.

Table 2 shows that this condition is not fully met. Customer Satisfaction is the most critical KPI. At Airbus, delivering helicopters on time is a top priority. The model must predict satisfaction accurately to support this objective.

This justifies a deeper analysis to understand where and why the model fails. The next section focuses on Customer Satisfaction and reformulates the problem as a classification task to better identify underperforming configurations. We do not apply the same approach to WIP and Investment, as their prediction performance is already acceptable.

5 CLASSIFICATION BASED-APPROACH:

In this section we reframe the problem from regression to classification to better align with industrial decision-making. We focus on scenarios where customer satisfaction falls below 0.90, as this situation risks contractual penalties and reputational damage.

5.1 Threshold Definition and Acceptance Ranges

To deepen this study, we engaged in discussions with simulation business experts at Airbus to better understand how these metrics directly impact production.

When customer satisfaction exceeds 90%, it indicates that the workshop parameters are well-optimized to ensure on-time helicopter delivery. Conversely, when customer satisfaction falls below 90%, it means the workshop configuration is insufficient to meet delivery deadlines.

Previously, model performance was evaluated using MAE and RMSE across the full range of customer satisfaction values, which penalized the model unfairly. Errors in regions where the satisfaction score is already low (e.g., predicting 20% instead of 80%) do not impact decision-making, since both values indicate a bad configuration. This led to a misleading assessment of model performance.

To address this, we introduce error tolerance thresholds based on the customer satisfaction range.

5.2 True vs. Predicted Classification Near Satisfaction Threshold

To better understand the model's behavior around the critical threshold of 0.90 for customer satisfaction, we provide two complementary plots. These visualizations help explain not just the magnitude of prediction errors, but also their decision impact.

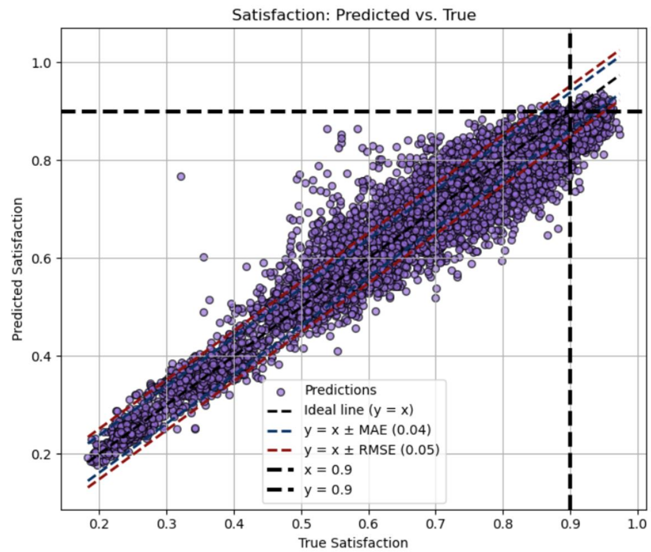

Figure 4: Prediction errors near the 0.90 threshold: tolerance zones based on MAE and RMSE.

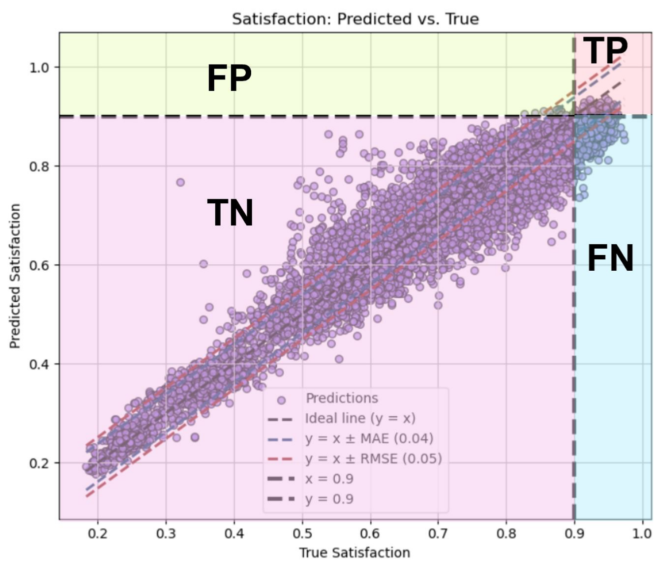

Figure 5: Classification of prediction outcomes: TP, TN, FP, FN zones.

Figure 4 shows the dashed lines for the ideal line ($y = x$), the MAE (±0.04), and RMSE (±0.05) bands, with the threshold at 0.90 marked in both axes. Most predictions fall within a narrow band around the true values, but a few outliers cross the decision threshold.

Figure 5 divides the prediction space into four regions:

- **True Positives (TP):** Correctly predicted satisfaction $\geq$ 0.90.
- **True Negatives (TN):** Correctly predicted satisfaction $<$ 0.90.
- **False Positives (FP):** Overestimated satisfaction ($\hat{y} \geq 0.90$, but $y < 0.90$).
- **False Negatives (FN):** Underestimated satisfaction ($\hat{y} < 0.90$, but $y \geq 0.90$).

This classification view supports a finer understanding of when the model produces incorrect decisions. Although the number of misclassified points is low, their position relative to the 0.90 threshold makes them critical in operational decision-making.

5.3 Classification Metrics

Accuracy measures the ratio of correct predictions to the total predictions. It uses true positives and true negatives. The formula divides the sum of true positives and true negatives by all cases. This metric shows overall performance.

Precision measures the correctness of positive predictions. It uses true positives and false positives. The formula divides true positives by the sum of true positives and false positives. This metric shows how often positive predictions are correct.

Recall measures the ability to identify all actual positive cases. It uses true positives and false negatives. The formula divides true positives by the sum of true positives and false negatives. This metric shows how many actual positives the model finds.

Each metric gives a different view of the model's performance.

We compute FP, FN, TN, and TP for our cases and then compute these metrics as follows :

Table 3: Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	TP = 2020	FN = 57
Actual Negative	FP = 48	TN = 6055

Accuracy

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} = \approx 0.9872$$

Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} = \approx 0.9768$$

Recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} = \approx 0.9726$$

The accuracy is 0.9872. This result shows that the model makes correct predictions in 8075 out of 8180 cases. The precision is 0.9768. This value means that when the model predicts a positive outcome, it is correct in 2020 out of 2068 cases. The recall is 0.9726. This value indicates that the model identifies 2020 out of 2077 actual positive cases. These metrics show that the model produces reliable results and has a low error rate.

5.4 Optimization with Genetic Algorithm and Surrogate Model

After training the MLP surrogate model on synthetic data generated through simulation, we integrated it into an optimization pipeline using a GA. This coupling allows the GA to evaluate a large number of workshop configurations rapidly, relying on the surrogate model instead of the original simulation.

The GA explores the search space by generating candidate configurations and uses the MLP model to predict the resulting KPIs (wip, investment and customer satisfaction). This process is significantly faster than running AnyLogic simulations for each candidate. As shown in figure 6, a single evaluation using the

MLP takes under a second, enabling the GA to converge toward high-performing solutions within a few minutes.

This approach combines the exploration capabilities of the GA with the speed of the surrogate model. It supports industrial use cases that require real-time or frequent decision-making. The results confirm that most optimized configurations satisfy operational constraints, particularly customer satisfaction thresholds.

	Model Design	Scenario Execution	Automated Execution	Target state : Real-Time Workshop Control
Exploration Method	Initial model development and calibration	Manual parameters adjustment	Genetic algorithm (crossover + mutation)	Genetic algorithm (crossover + mutation)
Evaluation Model for the Objective Function	AnyLogic model	Java version of the AnyLogic model	Java version +Genetic Algorithm	AI Model
Time for an Evaluation or a Prediction for a Set of Parameters	Not applicable	5 hours including data preparation and KPI calculation	15 minutes	Seconds
Average Number of Attempts	Not applicable	20	Between 140 and 200	Between 140 and 200
Time to Converge to a Solution	Not applicable	5 hours * 20 ≈ 4 days	15 minutes * 200 ≈ 2.08 days	Seconds




Figure 6: Calculation time for 20 simulation scenarios.

6 DISCUSSION:

This research introduces a classification-based evaluation framework to better align machine learning predictions with the operational priorities of Airbus Helicopters. In particular, Customer Satisfaction is the most critical KPI, as it reflects the ability to deliver helicopters on time. By introducing tolerance thresholds and reformulating the problem as a classification task, we were able to detect configurations that risk late delivery more clearly.

The classification metrics confirm the robustness of the model, with an accuracy of 98.72%, precision of 97.68%, and recall of 97.26%. However, the presence of 48 false positives (FP) and 57 false negatives (FN) indicates that some configurations near the satisfaction threshold remain difficult to classify correctly. False positives are especially risky, as they may lead to incorrect decisions and delivery delays.

The model performs well in the high satisfaction range (above 90%), but its reliability decreases near the decision boundary (80%–95%). This highlights the need to refine the model in these critical zones, where small errors may lead to wrong decisions.

Beyond prediction, the genetic algorithm was used to generate synthetic data. This exploration helped build a rich and varied dataset, allowing the surrogate model to generalize across many configurations. As shown in figure 6, once trained, the surrogate model provides results in under a second, making it possible to replace time-consuming simulations. This speed enables fast testing and real-time feedback.

Overall, this approach supports faster, more reliable industrial decisions. It avoids long simulation runs and provides safeguards by checking prediction reliability through classification. This reduces the risk of accepting configurations that may underperform. Future work could integrate ensemble methods or adaptive thresholds to improve the model's behavior around the decision boundary.

7 CONCLUSION

In this study, we presented a classification-based approach to evaluate the performance of an MLP model for predicting industrial KPIs at Airbus Helicopters. By reframing the problem from regression to classification, we were able to focus on the most critical aspects of decision-making, such as identifying configurations that risk late deliveries. The introduction of error tolerance thresholds and the use of precision, recall, and F1-score metrics provided a comprehensive evaluation of the model's performance.

The results demonstrate that the MLP model is highly accurate and reliable, with strong performance in predicting customer satisfaction, WIP, and investment. However, the presence of false positives and false negatives highlights the need for further refinement, particularly in the transition zone between acceptable and unacceptable configurations. By addressing these challenges, we can enhance the model's applicability and ensure that it supports real-time decision-making in complex industrial environments.

Our work contributes to the growing body of research on surrogate models for combinatorial optimization, demonstrating the potential of deep learning to replace traditional simulation models. The classification-based approach provides a structured framework for evaluating model performance, emphasizing industrial risk management over generic regression accuracy. Future research could explore the integration of additional data sources, such as real-time production data, to further improve the to further improve the model's predictive capabilities.

A APPENDICES

This table provides explanations of all abbreviations used in Figure 1, which illustrates the MECA 4.0 workshop architecture at Airbus Helicopters.

Table 4: Abbreviations and descriptions of elements in the MECA 4.0 workshop model.

Characteristics	Description
MGB	Main Gearbox
MRH	Main Rotor Head
MRO	Maintenance, Repair, and Overhaul
Serial	The manufacturing of new parts, produced in sequence as part of the continuous production process.
ZPP	A zone dedicated to the preparation of elementary parts before they are used in assembly operations.
CONWIP PSE	Elementary pre-stock for assembly of MRH and MGB
CONWIP MGB	Advance MGB stock
CONWIP MRH	Advance MRH stock
PSE Serial	Pre-stock assembly of MGB and MRH components
PSE MRO	Elementary parts of Maintenance, Repair, and Overhaul (MRO)
CA MGB	Refers to the assembly component of the Main Gearbox (MGB).
CA MRH	Refers to the assembly component of the Main Rotor Head (MRH).
Stock CA	The area where we store the assembly components of the Main Gearbox (MGB) and the Main Rotor Head (MRH)
FAL	Final Assembly Line, where the final integration of all helicopter components takes place before delivery.
Spare	Refers to extra parts kept in reserve to replace defective or worn-out components or to support maintenance activities.

REFERENCES

Ducoffe, M., G. Povéda, A. Galametz, R. Boumazouza, M.-C. Martin, J. Baris, *et al*. 2024. "Surrogate Neural Networks Local Stability for Aircraft Predictive Maintenance". *arXiv preprint arXiv:2401.06821*.

Forrester, A. I. J., A. Sobester, and A. J. Keane. 2008. *Engineering Design via Surrogate Modelling: A Practical Guide*. Chichester: John Wiley & Sons https://doi.org/10.1002/9780470770801.

Hürkamp, A., S. Gellrich, A. Dér, C. Herrmann, K. Dröder, and S. Thiede. 2021. "Machine learning and simulation-based surrogate modeling for improved process chain operation". *The International Journal of Advanced Manufacturing Technology* 117:2709–2724 https://doi.org/10.1007/s00170-021-07084-5.

Jin, Y. 2011. "Surrogate-assisted evolutionary computation: Recent advances and future challenges". *Swarm and Evolutionary Computation* 1(2):61–70 https://doi.org/10.1016/j.swevo.2011.05.001.

Li, F., Y. Li, X. Cai, and L. Gao. 2022. "A Surrogate-Assisted Hybrid Swarm Optimization Algorithm for High-Dimensional Computationally Expensive Problems". *Swarm and Evolutionary Computation* 72:101096 https://doi.org/10.1016/j.swevo.2022.101096.

Montgomery, D. C. 2017. *Design and Analysis of Experiments*. 9th ed. Hoboken, NJ: Wiley.

Paape, N., J. A. W. M. van Eekelen, and M. A. Reniers. 2024. "Simulation-based optimization of a production system topology: a neural network-assisted genetic algorithm". *arXiv preprint arXiv:2402.01511* 0(0).

Pereira, D., F. Afonso, and F. Lau. 2025. "End-to-End Deep-Learning-Based Surrogate Modeling for Supersonic Airfoil Shape Optimization". *Aerospace* 12(5):389 https://doi.org/10.3390/aerospace12050389.

Pestourie, R., Y. Mroueh, T. V. Nguyen, P. Das, and S. G. Johnson. 2020. "Active Learning of Deep Surrogates for PDEs: Application to Metasurface Design". *npj Computational Materials* 6(1):164 https://doi.org/10.1038/s41524-020-00431-2.

Saadi, M., V. Bernier, G. Zacharewicz, and N. Daclin. 2024, May. "Surrogate Modelling with Deep Learning for Optimizing Manufacturing Assembly Lines". In *Proceedings of the 2024 Annual Modeling and Simulation Conference (ANNSIM'24)*, 1–12. Washington, DC, USA: The Society for Modeling & Simulation International (SCS): SCS https://doi.org/10.23919/ANNSIM61499.2024.10732237.

Saves, P., R. Lafage, N. Bartoli, Y. Diouane, J. Bussemaker *et al*. 2024. "SMT 2.0: A Surrogate Modeling Toolbox with a focus on hierarchical and mixed variables Gaussian processes". *Advances in Engineering Software* 188:103571 https://doi.org/10.1016/j.advengsoft.2023.103571.

Stander, L., M. Woolway, and T. van Zyl. 2022, 05. "Surrogate-assisted evolutionary multi-objective optimisation applied to a pressure swing adsorption system". *Neural Computing and Applications* 37:739–755 https://doi.org/10.1007/s00521-022-07295-1.

Suarez, G., E. Özkaya, N. Gauger, H.-J. Steiner, M. Schäfer, and D. Naumann. 2024. "Nonlinear surrogate model design for aerodynamic dataset generation based on artificial neural networks". *Aerospace* 11(8):607 https://doi.org/10.3390/aerospace11080607.

AUTHOR BIOGRAPHIES

MARYAM SAADI is a PhD candidate at SyCoIA, IMT Mines Ales, France, in collaboration with Airbus Helicopters, Marignane, France. She holds a Master's degree in Industry 4.0 from Mohammed VI Polytechnic University, Morocco. Her research focuses on artificial intelligence and simulation applied to industrial workflows at Airbus. Her email address is maryam.saadi@airbus.com.

VINCENT BERNIER is an Industrial Model Based System Engineering Expert at Airbus Helicopters in Marignane, France. Previously, he worked as an Engineer and Researcher at Peugeot in Poissy, France. He received his PhD at the Graduate School of Industrial Engineering, in Grenoble, France. His thesis focused on a new flow management policy: re-sequencable scheduling. His email address is vincent.bernier@airbus.com.

GRÉGORY ZACHAREWICZ is a Full Professor at *SyCoIA, IMT Mines Ales, France*. Previously, he was an Associate Professor (HDR) at the University of Bordeaux, France. He received his PhD from Paul Cézanne University, Aix-Marseille III, France. His research focuses on distributed modeling and workflow simulation. His email address is gregory.zacharewicz@mines-ales.fr.

NICOLAS DACLIN is a Full Professor at *SyCoIA, IMT Mines Ales, France*. Previously, he was a Research and Development Engineer at ARMINES in Nîmes, France. He received his PhD from the University of Bordeaux, France. His research focuses on enterprise modeling, process control, and automation. His email address is nicolas.daclin@mines-ales.fr.