The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-025-16070-0

ORIGINAL ARTICLE

Modeling of tensile strength and wear resistance in friction stir processed MMCs by metaheuristic optimization and supervised learning

Vahid Modanloo[1] · **Majid Elyasi**[2] · **Taeyong Lee**[3] · **Luca Quagliato**[3]

Received: 2 May 2025 / Accepted: 2 July 2025

Abstract

This study presents a systematic framework for modeling the tensile strength (TS) and wear resistance (WR) of friction stir processed (FSP) AA1100 metal matrix composites (MMCs) reinforced with aluminum oxide (Al_2O_3) powder reinforcement. Using a design of experiments approach, the influence of tool rotation speed, feed rate, and reinforcement content on TS and WR was investigated across 20 independent experiments, with the optimal process combination resulting in a 20.7% increase in TS and a minimum WR of 7.3×10^{-3} mm^3/m. To capture the nonlinear relationships between process parameters and material performance, TS and WR were initially modeled using second-order polynomial regression, further optimized by two metaheuristic (ME) algorithms. Furthermore, two supervised machine learning (ML) models were developed, optimized, and benchmarked. ME and ML models were validated using both the experimental dataset and ten synthetic cases generated by Gaussian mixture models (GMM) within the experimental features' latent space. For TS modeling, ME-optimized regressions showed the mean percentage deviations in 15.6% ~ 27.7% range while the ML formulations in the 12.4% ~ 17.9% range. For WR, a similar reduction in error range and variance was also confirmed, highlighting a higher modeling reliability than the TS counterpart. Results demonstrate that while both ME and ML approaches can be employed for FSP process modeling within the latent space, data-driven approaches, and especially neural networks architecture, have a clear advantage in prediction accuracy and robustness under data-limited conditions. Although modeling with only 20 cases is challenging for both ML and ME methods, such data scarcity is typical in industrial settings; thus, this study reflects a realistic scenario where reliable predictions must be achieved from limited data. Overall, this work provides a practical modeling benchmark between ME-optimized and ML modeling for the FSP and may help in identifying the best modeling approach also for other manufacturing engineering processes.

Keywords Friction stir processing (FSP) · Al/Al_2O_3 composite · Metaheuristic model · Supervised machine learning · Process setting and optimization

1 Introduction

Aluminum-based metal matrix composite (MMC) materials are a growing phenomenon in the automotive and aerospace industries thanks to their high corrosion and wear resistance, specific moduli, and high strength-to-weight ratio [1–3]. Among various manufacturing processes, a relatively new and viable option to realize MMC is represented by friction stir processing (FSP) [4], a solid-state welding process in which a non-consumable rotating tool is applied to the workpiece surface, generating localized plastic deformation and stirring the material as the tool moves linearly. Apart from being a welding technique, the FSP is also employed to promote the joining of reinforcement particles to a metal

Vahid Modanloo and Luca Quagliato contributed equally to this work.

✉ Vahid Modanloo
v.modanloo@sirjantech.ac.ir

✉ Luca Quagliato
lucaq@ewha.ac.kr

[1] Mechanical Engineering Department, Sirjan University of Technology, Sirjan, Iran

[2] Faculty of Mechanical Engineering, Babol Noshirvani University of Technology, Babol, Iran

[3] Department of Mechanical and Biomedical Engineering, Ewha Womans University, Seoul, South Korea

Published online: 14 July 2025

substrate to manufacture MMC [5, 6]. Considering the widespread application of aluminum products, most of the research effort is concentrated on aluminum-based matrices, combined with various oxides and carbide particles, to improve key properties by creating MMC, as hereafter summarized.

In Sahraeinejad et al. [5], Al_2O_3 (aluminum oxide), SiC (silicium carbide), and B_4C (boron carbon) particles were added to an aluminum 5059 (AA5059) matrix by FSP resulting in a 32% increase in the yield strength for the case of 4.2 μm particle size. In their research, the smaller particle size allowed for a higher filling ratio, also associated with the highest improvement in mechanical properties. Similarly, the investigation of the effect of SiC particles on the microstructure and mechanical properties of AA6061-based MMC using a multi-pass FSP showed strength improvements from 16.1 to 72.3 MPa for the first and sixth pass, respectively, for a baseline of 275.7 MPa, demonstrating the progressive strengthening [7]. Similar improvements were also confirmed for the tensile and compression properties for the case of AA7075 matrix reinforced by hexagonal boron nitride (HBN), with improvements equal to 26.5% and 40%, respectively [8]. In addition to the reinforcement effect, the FSP is highly influenced by the process parameters employed and mainly by rotational speed, feed rate, applied pressure, and tool angle [9]. On top of that, both reinforcement material and FSP parameters' optimization are linked to one another, and both influence the mechanical performance of the FSPed component [10]. In this regard, if a non-optimized combination of FSP parameters and reinforcement material is employed, the risk for intermetallic and complex phase generation arises, leading to general embrittlement and weakening rather than reinforcement [11].

When material and processing aspects are considered together in the optimization, in addition to promoting a strong bond between matrix and reinforcement, the FSP process also showed to mitigate Tungsten inert gas (TIG) welding defects such as coarse grain structure, porosity, microvoids, and solidification cracking while also refining the microstructure of both dendrites and intermetallic phases [12]. In a similar way, the effect of the reinforcement weight percentage (wt%) on microstructure, mechanical properties, and wear resistance was investigated by Thangarasu et al. [13] for the case of a TiC-reinforced AA6082 matrix and a fixed travel speed of 60 mm/min. From their results, the increase in TiC wt% was linked to higher strength, a decrease in ductility, and a reduction in the area subjected to plastic deformation in the matrix thanks to a reduction in the stir zone. In terms of a comprehensive study, Devaraju et al. [14] synthesized aluminum surface composites reinforced with SiC and Al_2O_3 particles using the FSP. By varying the rotational speed and reinforcement wt%, they investigated the effect on tensile strength, hardness, and wear rate showing improvements in all categories as well as a reduction in the particle size from the original one.

In addition to the literature state of the art summarized so far, other contributions dealing with FSP-based MMC are relevant to post-production reinforcement distribution and wear properties [15], grain size refinement [16], mechanical properties and corrosion resistance [17], and tribological properties [18]. It is worth noting that, thanks to its widespread use, most of the work is carried out by employing aluminum as the matrix material, although the typology of the reinforcement varies according to the application field and targeted improvement [19]. Nevertheless, MMC manufacturing by FSP does not refer only to aluminum-based matrices, but also to other metallic materials; among them are copper [20, 21], steel [22], and titanium alloys [23].

In addition to the large variety of works relevant to the FSP process, in recent years, the friction stir vibration processing (FSVP) has emerged as a promising modification to standard FSP, incorporating mechanical vibration to further enhance microstructure and mechanical properties in comparison to the standard FSP. In this regard, Bagheri et al. [24] performed a direct comparison between FSP and FSVP for the development of magnesium surface nanocomposites, reporting notable improvements in grain refinement, ultimate tensile strength (UTS), and formability with the addition of vibration to the FSP. In a similar way, Abbasi et al. [25] demonstrated that applying vibration during FSP on Al5052 alloy led to a 33% reduction in grain size, a 7% increase in UTS, and an increase in dislocation's density thanks to an enhanced dynamic recrystallization. Same concerning aluminum-based MMC, Barati et al. [26] found that thanks to the achievement of finer grains and a more uniform particle distribution, FSVP significantly improved strength, ductility, wear resistance, and corrosion properties in Al6061/SiO2 MMC in comparison to the standard FSP. Similar results were also obtained by Abbasi et al. [27] in relation to Al5052/SiC nanocomposites and linked to the increased strain granted by applying vibrations in the FSP process, a fact promoting dynamic recovery and recrystallization. More recently, Abdollahzadeh et al. [28] and Bagheri et al. [29] systematically investigated AZ91/SiC composites, showing that FSVP enhances strength, wear resistance, hardness, and ductility. In addition to that, it ought to be pointed out that, while FSVP yields better performance than FSP, the improvement in microstructure intensifies as vibration frequency or the number of passes increased [29]. On top of that, a more recent contribution by Men et al. [30] investigated the microstructural, mechanical, and tribological properties of AZ31/AlFeCrMoNb surface composite manufactured by FSVP and found that employing vibrations increases hardness by 20% and shear strength by 33%, respectively.

Despite these substantial advances in FSP and FSVP processing technologies, as summarized so far, controlling and optimizing the mechanical and tribological performance of the fabricated MMCs remains highly dependent on the precise selection and tuning of process parameters. In this regard, considering the rapid development of machine learning (ML) techniques in the field of material science and engineering [31], including additive manufacturing [32] and metal forming [33], recent contributions demonstrated and summarized the large use of ML models in friction stir welding [34–36], although no contributions seem to deal with an in-depth analysis of the possibility and performance of data-driven modeling approaches for the FSP and FSVP.

However, two recent contributions, summarized hereafter, show the possibility of employing numerical or intelligent modeling approaches to analyze the complex interaction between process and material parameters and the resulting mechanical performance of FSPed components. The former contribution refers to the alloying of a base AA1050 with Cu powder through FSP experiments, followed by training of a genetic programming algorithm, subsequently employed to predict and optimize the material properties of the MMC [37]. The latter instead considered a universal differential equation (UDE) frame and constructed a data-driven model to reduce dynamics to lumped parameters, reducing complexity while allowing for the link between tool power input and temperature measurement for process control [38]. Though the former contribution is based on a machine learning model, the latter combines machine learning, in terms of neural-based ordinary differential equations (ODEs), with lumped parameter methods, making it closer to an approximation of a metaheuristic method, since it infers the system dynamics using data-driven modeling.

As summarized so far, despite various research works focused on investigating the FSP and FSVP in terms of reinforcement materials' influence, mechanical properties, microstructure, and tribology [39], and the recent effort to propose data-driven modeling techniques [37, 38], a comprehensive study on the best modeling methodology for correlating the FSP, or FSVP, process parameters with the key output variables, such as strength and wear resistance improvements, seems to be missing in the literature. As demonstrated in the recent literature [37, 38], both machine learning and metaheuristic approaches seem to be promising in modeling the features to target variables relationship in the FSP process, but no studies have highlighted whether one approach is superior to the other and whether this superiority is subordinated to the boundaries of the considered problem.

To address this gap, this research is subdivided into three main areas, which also represent its key novelty points. First, the present work systematically benchmarks two regression models, optimized by the Differential Evolution [40, 41] and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [42, 43] metaheuristic (ME) algorithms, and two supervised machine learning (ML) models, Extreme Gradient Boosting (XGB) [44] and Deep Neural Network (DNN) [45]. To this end, this research employs an experimentally derived dataset of 20 FSP-fabricated AA1100-Al_2O_3 MMCs to explore the influence of rotation speed, feed rate, and reinforcement weight on the tensile strength and wear rate. TS was measured by tensile tests on specimens having the FSPed zone located in their center, whereas WR by linear wear experiments on the FSPed zone by measuring the amount of material loss. Second, this research demonstrates the implementation procedure, advantages, and limitations of ME-optimized regression (by DE and CMA-ES) and supervised ML (XGB and DNN) models with a special focus on the challenges posed by employing a small dataset, a common case for industrial manufacturing scenarios. In this regard, both approaches are trained and validated using the same 20 experimental cases, and their interpolation accuracy and robustness were further assessed on synthetic data generated by tuning a Gaussian mixture model (GMM) model [46, 47], for a total of 10 test cases belonging to the same distribution of the experimental feature space. Third, the experimental results are presented and commented upon considering the established knowledge of the FSP process and provide a critical analysis of the suitability, reliability, and practical limitations of each modeling approach under limited data conditions. By doing so, the performance of ME and ML models on both experimental and synthetic data provides practical and useful insights for model selection, design, tuning, and optimization for the FSP process. The overall research implementation is depicted in Fig. 1, which also serves as the backbone for navigating the content reported in the "Materials and methods" and "Results" sections.

In summary, this work provides a direct and systematic analysis of ME-optimized regression and supervised ML models for predicting tensile strength (TS) and wear resistance (WR) in friction stir processed AA1100-Al_2O_3 MMCs under limited data conditions. The prediction accuracy, robustness, and practical applicability of each approach are critically evaluated by employing both experimental and GMM-generated data, providing a clear and fair analysis of their interpolation performance. In this regard, and from the results of this research, ME-optimized regressions showed the mean percentage deviations ranging from 15.6% to 27.7%, while the ML formulations went from 12.4% to 17.9%, respectively. Similar trends, although with lower magnitudes, were also found for WR. These findings demonstrate that, while both ME and ML approaches are viable modeling options for the FSP process, data-driven ML models, and especially the DNN formulation, consistently deliver higher accuracy and reliability under limited data conditions. This comprehensive analysis reveals the

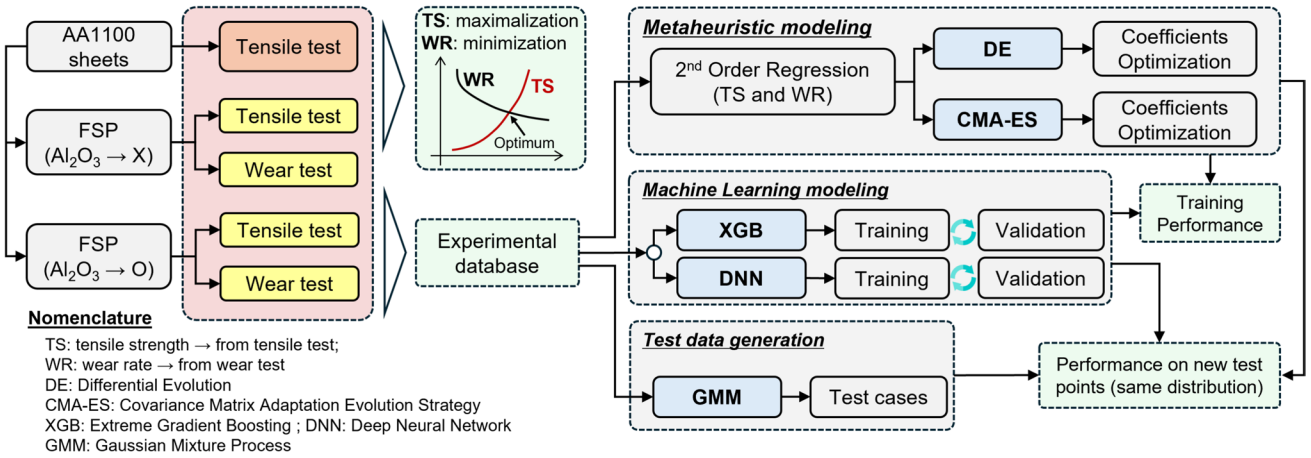


Fig. 1 Overview of the research content with interactions among the phases, as summarized within the "Materials and Methods" and "Results" sections

strengths and limitations of ME and ML strategies in modeling the FSP process on a small-scale dataset. Moreover, the analysis, results, and considerations provided throughout the manuscript offer guidance for engineers and researchers seeking robust predictive modeling solutions in similar manufacturing scenarios. Ultimately, this study advances the state of the art by establishing a transparent, reproducible framework for process-property modeling in FSP/MMCs, supporting the model selection phase and the concurrent process-property optimization.

2 Materials and methods

2.1 Material characterization

The base material employed in this research is an AA1100 sheet with a 2-mm thickness and chemical composition reported in Table 1. To assess the mechanical properties, tensile tests based on the ASTM-E8 standard were carried out by wire-cutting the specimens from the aluminum sheet and testing them on a SANTAM tensile testing machine with a 30-ton capacity. The base AA1100 properties are reported in Section 2.2.

2.2 FSP process and characterization

To perform the FSP experiments, commercially available 2-mm thickness AA1100 sheets were initially wire-cut into dimensions of 100×150 mm. A schematic representation of the FSP process is reported in Fig. 2a. In the FSP experiments, a clamping system was employed to fix the part into position and avoid unwanted movement and excessive vibrations, as shown in Fig. 2b. A single-pass FSP was carried out by employing a tool with a shoulder diameter of 20 mm ending with a cube-shaped pin with a 6 mm side and a 1.8 mm height, as shown in the detail of Fig. 2b, until the full length of 150 mm was completely friction stir processed (Fig. 2c).

For each experiment involving Al_2O_3 reinforcement, the FSP area was filled with a uniform layer of Al_2O_3 powder with an average diameter of 30 µm and 3.6 g/cm^3 density. The total weight of the particles was measured beforehand, but the particles' distribution on the top of the AA1100 sheet in the FSP area was not included in the experimental setup. The FSP experiments were carried out considering three levels for the rotational speed of the tool (N), three feed rates (f), and three Al_2O_3 powder weights (W), as summarized in Table 2, for a total of 20 experimental combinations, as reported in Table 3, each one representing the average of three repetition tests with the same combination of parameters. The 20 combinations of Table 3 were determined through the central composite design (CCD) statistical technique, considering 3 parameters and 3 levels of configuration. The three selected levels for each parameter are based on the feasible settings of the experimental setup, preliminary investigations by the authors, and the established parameter ranges summarized in recent literature reviews [6].

After the completion of each FSP experiment, two characterization tests were carried out to investigate the resulting performance of the manufactured MMC. To this end, tensile test specimens according to the ASTM-E8 standard

Table 1 Chemical composition of the used AA1100 sheet

Element	Al	Mn	Mg	Cu	Fe	Si	Mg
Percentage (%)	Bal	0.004	0.004	0.061	0.508	0.15	0.9

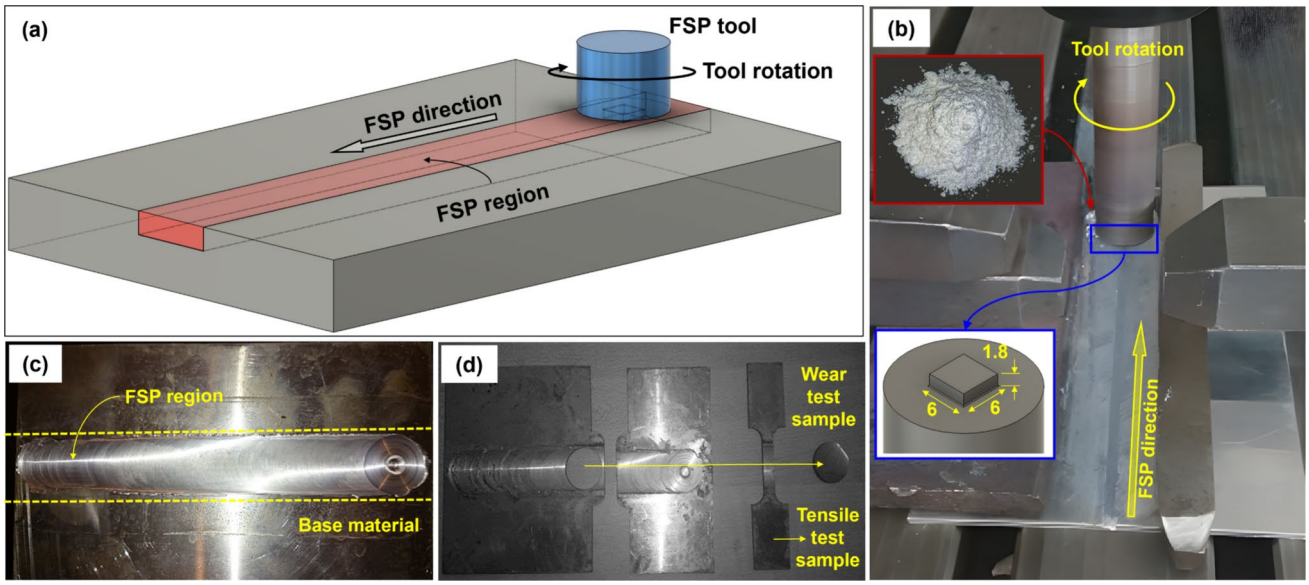


Fig. 2 **a** Schematic representation of the FSP process. **b** FSP process on AA1100 sheet including details of the processing tool geometry and Al_2O_3 powder reinforcement and **c** AA1100 sheet after the FSP process. **d** Extraction of tensile specimens from the FSPed plate to assess the tensile strength (TS) and circular specimen cut from the FSPed plate after the wear test to assess the wear rate (WR)

Table 2 FSP process parameters and relevant levels

Parameter	Symbol	Unit	Levels		
Tool rotary speed	N	rpm	800	1200	1600
Feed rate	f	mm/min	40	60	80
Al_2O_3 powder	W	g	0	0.3	0.6

were cut by placing the FSPed region right in the middle of the calibrated zone, as shown in Fig. 2d. In light of the close-to-one aspect ratio of the Al_2O_3 powder, no anisotropic behavior was observed, and for the specimens perpendicular to the FSP direction, the grip part was limited to 5 mm on each side of the base material, due to the width of the FSP region. The true stress–strain curves for base AA1100 material, FSP without, and with Al_2O_3 powder (maximum improvement case), are reported in Fig. 3, whereas the summary of the tensile strength (TS) recorded from the tensile tests is reported in Table 3.

In addition to that, wear tests were carried out on the FSP region (Fig. 2d) by employing a Φ25 mm flat pin head, a 10 N normal load, 3.0 m/s linear speed, and a sliding distance of 100 m. The weight of Φ25 circular specimens, cut from the wear test region, was measured after the tests with a digital weighing balance with 0.0001 g accuracy. After each wear test, the wear rate (WR) was determined as the weight loss from the specimen in comparison to a non-FSPed specimen of the same size. The recorded WR for each one of the experiments is reported in Table 3 together with the experimental settings and the TS results.

Table 3 FSP experiments set-up with TS and WR results

#	*N* (rpm)	*f* (mm/min)	*W* (g)	TS (MPa)	WR·10^{-3} (mm^3/m)	#	N (rpm)	f (mm/min)	W (g)	TS (MPa)	WR·10^{-3} (mm^3/m)
1	800	40	0	256	8.2	**11**	1200	40	0.3	396	7.3
2	1600	40	0	294	6.8	**12**	1200	80	0.3	222	7.8
3	800	80	0	196	9.7	**13**	1200	60	0	220	8.1
4	1600	80	0	213	7.8	**14**	1200	60	0.6	371	7
5	800	40	0.6	337	7.7	**15**	1200	60	0.3	324	8
6	1600	40	0.6	371	7.8	**16**	1200	60	0.3	320	7.8
7	800	80	0.6	311	7.4	**17**	1200	60	0.3	318	7.9
8	1600	80	0.6	341	7	**18**	1200	60	0.3	328	8.1
9	800	60	0.3	196	9.1	**19**	1200	60	0.3	329	8
10	1600	60	0.3	341	8.2	**20**	1200	60	0.3	324	7.3

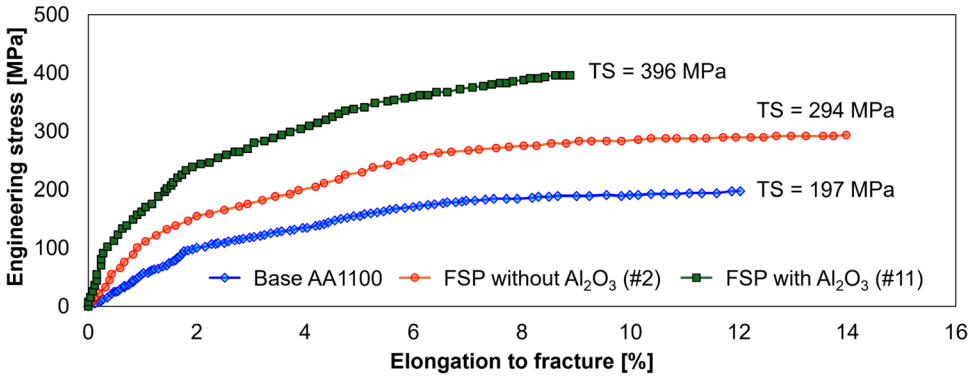


Fig. 3 Engineering stress–strain curves for base AA1100, FSP without, and with Al_2O_3 powder reinforcement

2.3 Polynomial regression and metaheuristic (ME) models

Meta-heuristic (ME) methods are employed to address complex problems where identifying the optimal solution is challenging or impractical by providing an optimal solution in local boundaries rather than global optimization. Considering the scope of this research to predict TS and WR based on three features, the Differential Evolution (DE) algorithm [40, 41] and the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [42, 43] have been considered and combined with tuning of their hyperparameters based on the *GridSearch* method. Both ME algorithms have been applied for the optimization of the constants of the regression functions for TS and WR to minimize the deviations between predicted and true values.

First, considering that all features and target variables have different scales of magnitude, the *Min–Max* scaling was applied to the dataset of Table 3. In addition to that, the coefficients' optimization discussed in this section and carried out via the two ME models, refers to $[\theta_1,\ldots,\theta_9]$ as in Eq. (1) resulting from a standard regression based on the analysis of variance (ANOVA) analysis.

$$y = \theta_0 \cdot 1 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 {X_1}^2 + \theta_5 {X_2}^2 + \theta_6 {X_3}^2 + \theta_7 X_1 X_2 + \theta_8 X_1 X_3 + \theta_9 X_2 X_3$$
$$X_{poly} = \left[1, X_1, X_2, X_3, {X_1}^2, {X_2}^2, {X_3}^2, X_1 X_2, X_1 X_3, X_2 X_3\right] \quad (1)$$
$$where: X_1 = N, X_2 = f, X_3 = W$$

In Eq. (1), y represents the generalized regression function, with the relevant θ coefficients, whereas X_{poly} is the normalized features' vector. Considering the size of the experimental database, composed of the 20 cases of Table 3, the choice of employing a second-order polynomial regression structured as (1), and not including higher-order terms, is motivated by three reasons. First, a third-order polynomial function, with its 20 coefficients, would result in the exact passing of the interpolation function through all the data points, leaving few to no space for optimization by the ME models. In a similar way, the utilization of a second-order function allows for a lower constraint in the predictions in between data points; in other words, employing a lower-order function suits better the aims of this research of describing the whole features' space and not only the 20 data points themselves. Finally, employing up to second-order terms allows for a better interpretation of the physical role of each one of them, as presented in Sect. 3.1 of the paper. For all these reasons, in this research, a second-order polynomial function was employed for the definition of the base regression models for TS and WR.

For the optimization of the coefficients, the mean absolute percentage error (MAPE), Eq. (2), with respect to the whole dataset was employed. As a remark, TS and WR were modeled separately but considering the same regression function form of (1) and the deviations estimated by the same MAPE error formulation of (2).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - f\left(X_{poly};\theta\right)}{y_i}\right|[\%] \quad (2)$$

In Eq. (2), y_i represents the true values for TS and WR, whereas $f(X_{poly};\theta)$ represents the predicted value using polynomial regression for a given set of θ coefficients. To this end, the aforementioned coefficient optimization is represented by the identification of the best set of θ coefficients that minimizes the MAPE.

The first of the two ME approaches employed to optimize the θ coefficients is the DE model, which consists of 3 phases: *mutation*, *crossover*, and *selection*. The DE equations frame summarized here are based on the one presented in Lu et al. [40] and Salgotra and Gandomi [41]. The initialization for the population, which refers to the θ coefficients, is based on a random (*rand*) value within the user-defined boundaries, as in Eq. (3). Hence, the population size (*pop-size*) is a hyperparameter defined in terms of a multiplication

factor for the coefficients' vector and is optimized during *GridSearch*.

$$\theta_i^{(0)} = \theta_{\min} + rand\left(\theta_{\max} - \theta_{\min}\right) 0 \le rand \le 1 \in \mathbb{R} \tag{3}$$

As concerns the *mutation* phase, each population member generates a mutation vector $v_i^{(g)}$, as in Eq. (4), defined for the *i*-individual and the *g*-generation. In Eq. (4), based on the *best/1/bin* strategy [44], $x_{rand\_1}^{(g)}$, $x_{rand\_2}^{(g)}$, and $x_{rand\_3}^{(g)}$ are three not equal and randomly selected individuals from the population, whereas *F* is the mutation factor, also considered as a hyperparameter in the *GridSearch*. This *mutation* phase aims to introduce diversity and explore different combinations of the original population in the search for a better global *minima* of the function.

$$v_i^{(g)} = x_{rand\_1}^{(g)} + F \cdot \left(x_{rand\_2}^{(g)} - x_{rand\_3}^{(g)}\right) \tag{4}$$

After each individual in the population is generated, a uniform *crossover* is applied to combine mutant vectors ($v_{i,j}^{(g)}$) with current population vectors ($x_{i,j}^{(g)}$) to create trial vectors ($u_{i,j}^{(g)}$), as defined in Eq. (5), where j_{rand} is introduced to have at least one mutation to be considered as a trial vector.

$$u_{i,j}^{(g)} = \begin{cases} v_{i,j}^{(g)} \text{ if } rand_j \le CR \cup j = j_{rand} \\ x_{i,j}^{(g)} \text{ otherwise} \end{cases} \\ 0 \le rand_j \le 1 \in R \text{ and } j_{rand} \in N \tag{5}$$

Trial vectors are, for all means and purposes, new candidates for the θ coefficients, evaluated by the crossover probability (CR), also defined as a hyperparameter. Considering that $CR \in [0,1]$, a value < 0.5 is defined as low and leads to a conservative choice where few mutations are introduced. On the other hand, a value of $CR \approx 1$ results in more exploration, thus prioritizing mutations over original populations' individuals.

The final step of each iteration in the DE optimization is represented by the *selection* phase, where the newly created trial vector ($u_{i,j}^{(g)}$), defined during the crossover, is evaluated and selected or rejected for the next generation, as summarized in Eq. (6), and is based on the MAPE evaluation criteria of Eq. (2).

$$x_{i,j}^{(g+1)} = \begin{cases} u_{i,j}^{(g)} \text{ if } MAPE\left(u_{i,j}^{(g)}\right) \le MAPE\left(x_{i,j}^{(g)}\right) \\ x_{i,j}^{(g)} \text{ otherwise} \end{cases} \tag{6}$$

As a final remark, the best mutation factor (*F*), crossover probability (*CR*), population size (*N*), and maximum number of iterations (*max_iter*) were all considered hyperparameters and are further discussed in Section 2.5.

The second employed ME model is the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a derivative-free approach particularly effective in solving non-linear, non-convex, and high-dimensional optimization functions [48]. CMA-ES optimizes the objective function, previously defined as in Eq. (1), by iteratively updating a multivariate normal distribution, from which candidate solutions are sampled [49], by adapting the mean vector and the covariance matrix. As with the DE algorithm, the CMA-ES is also divided into phases, namely, *sampling*, *evaluation*, *selection*, and *adaptation*. The equations framework for the CMA-ES model presented hereafter is based on recent literature contributions as implemented in Hansen [49].

In the *sampling* phase, at each *t*-iteration, a set of λ candidates is generated from a multivariate distribution, as in Eq. (7), where $x_k^{(t)}$ is the sampled solution, $m^{(t)}$ is the mean vector of the current distribution, $\sigma^{(t)}$ controls the step size of the global variance, $C^{(t)}$ is the covariance matrix, and $N\left(0, C^{(t)}\right)$ represents a multivariate distribution with a mean value equal to zero and covariance $C^{(t)}$.

$$x_k^{(t)} = m^{(t)} + \sigma^{(t)} \cdot N\left(0, C^{(t)}\right), k = 1, ..., \lambda \tag{7}$$

Afterwards, in the *evaluation* phase, each solution candidate ($x_k^{(t)}$) is evaluated on the objective function (1), allowing for the calculation of the corresponding fitness score, as in Eq. (8), where f_k is the fitness value of the *k*-th solution.

$$f_k = f\left(x_k^{(t)}\right), k = 1, ..., \lambda \tag{8}$$

The best solutions (μ) from the evaluation of (8) are identified in the *selection* phase by following a ranking-based approach where solutions with lower objective function values are preferred, as in Eq. (9).

$$x_{i:\lambda}^{(t)} = \text{argsort}_i\left(f\left(x_k^{(t)}\right)\right), k = 1, ..., \lambda \tag{9}$$

After selecting the top performing solutions ($x_{i:\lambda}^{(t)}$), the mean vector is updated from the *t*-step to the $t+1$ step as the weighted sum of these solutions, as in Eq. (10) where w_i are the selected weights.

$$m^{(t+1)} = \sum_{i=1}^{\mu} w_i x_{i:\lambda}^{(t)} \tag{10}$$

The final step of each iteration in the CMA-ES pipeline is represented by the update of the covariance matrix by the cumulative natural gradient adaptation principle, as in Eq. (11), where c_c is the learning rate controlling the covariance adaptation.

$$C^{(t+1)} = \left(1 - c_c\right) C^{(t)} + c_c \sum_{i=1}^{\mu} w_i \left(x_{i:\lambda}^{(t)} - m^{(t)}\right) \left(x_{i:\lambda}^{(t)} - m^{(t)}\right)^T \tag{11}$$

For the CMA-ES model, the evaluation metric is based on the MAPE, Eq. (2), and was employed to identify the best set of θ coefficients that minimize the objective function.

2.4 Machine learning (ML) modeling

To propose a modeling alternative and benchmark the performance of the ME model presented in the previous section, two highly regarded machine learning (ML) models were considered, namely the Extreme Gradient Boosting (XGB) [44] and Deep Neural Network (DNN) [45].

In the XGB model, based on the structure presented in Lee et al. [44], the objective function combines the loss function and the regularization term to optimize the model, Eq. (12), where the first term measures the difference between the predicted and true values, whereas the second part (regularization term) penalizes the model complexity.

$$L(\phi) = \sum_{i=1}^{n} \ell\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{T} \Omega\left(f_k\right) where \Omega\left(f_k\right) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2 \tag{12}$$

In Eq. (12), y_i are the true values, $\hat{y}_i$ the predictions, T is the number of leaves in a tree, γ controls the complexity of the model, λ regularizes the leaf weights, and ω are the leaf weights. The XGB model builds the model iteratively by adding one tree at a time to minimize the objective function, as in Eq. (13), where $f_t\left(x_i\right)$ is the new tree at the t-iteration. To determine the structure of the tree at each iteration, the objective function is approximated using a second-order Taylor expansion, as in Eq. (14) where g_i and h_i are the first- and second-order gradients.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t\left(x_i\right) \tag{13}$$

$$L^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2\right] + \Omega\left(f_t\right) \tag{14}$$

To maximize the prediction accuracy, TS and WR have been predicted by separate XGB models, each one optimized with its specific hyperparameters through a $k = 5$-*fold* cross-validation procedure.

The DNN model was implemented according to the pipeline described in Modanloo et al. [45], and its architecture is shown in Eq. (15), where the ReLU (rectified linear unit) activation function was employed and x represents the input features' vector. ReLU introduces non-linearity into the network, enabling it to learn complex patterns, and mitigates the vanishing gradient problem observed with Sigmoid and Tanh activations.

$$h_i = \mathrm{ReLU}\left(W_i x + b_i\right) where x = \left[x_1, x_2, ..., x_n\right] \in \mathfrak{R}^n \tag{15}$$

For each epoch, the loss function is estimated according to the mean squared error (MSE), as in Eq. (16), and employed in the backpropagation to update weights (W_i) and biases (b_i), summarized in Eq. (17) as ε, through an adaptive moment estimation (Adam) optimizer with a learning rate η.

$$L\left(y, \hat{y}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{16}$$

$$\frac{\partial L}{\partial \varepsilon} = -\frac{2}{N}\left[\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)\right]\frac{\partial \hat{y}_i}{\partial \varepsilon} where \varepsilon_{t+1} = \varepsilon_t - \eta\frac{\partial L}{\partial \varepsilon} \tag{17}$$

In this research, the MSE was chosen as the loss function as it penalizes more severely large residuals, ensuring the model minimizes significant errors effectively in a short number of epochs. In addition to that, the MSE is a differentiable function, making it easy to apply in the backpropagation of a DNN. On the other hand, for the estimation of the prediction accuracy, for both XGB and DNN models, the MAPE formulation of Eq. (2) was employed, allowing for a fair comparison with the residuals associated with the ME-optimized regressions.

2.5 Database pre-evaluations and models' training

First, the results of tensile and wear tests, in terms of TS and WR, were considered as inputs for an analysis of variance (ANOVA) to determine the adjusted sum of squares (Adj SS), F-value, and P-value. The Adj SS represents the portion of the total variability in the data that is explained by each factor after accounting for other factors in the model. The Adj MS, defined as the adjusted sum of squares (Adj SS) divided by the degrees of freedom (DoF), was not considered since all the variables have DoF = 1, making Adj SS = Adj MS. The F-value is used to assess whether the variation due to a factor is significantly higher than the residual error, whereas the P-value determines whether the factor is statistically significant. In this regard, a high F-value and a low (< 0.05) P-value are indicators of a strong and statistically significant influence on the response variable. The ANOVA analysis was carried out in the Python/Spyder 3.11 environment by employing *pandas*, *numpy*, *matplotlib*, and *statsmodels* libraries.

As concerns the ME models, the information relevant to the pre-processing and implementation steps is reported as follows. For the DE model, a *GridSearchCV* was employed with the following ranges to identify the best combination allowing for the minimizing of the MAPE: *mutation_factors*: [0.5, 0.7, 0.9], *recombination_rates*: [0.7, 0.9], *population_sizes*: [20, 50, 100], and *max_iterations*: [500, 1000, 2000]. In addition to that, the search space for the best set of θ coefficients is defined in a $[-500, 500]$ radius from their original point coming from the initial regression. In a similar way, the parameters for the CMA-ES model were optimized

using a *GridSearchCV* over *sigma*: [10, 20, 50, 100, 250] and *popsize:* [100, 250, 500]. For both ME models, the best θ coefficients were identified by minimizing the average MAPE over the whole of 20 cases (Table 3) and separately for TS and WR. The DE and CMA-ES codes were implemented in a Python/Spyder 3.11 environment considering *pandas*, *numpy*, *sklearn*, *scipy.optimize*, and *cma* libraries on the same workstation.

Concerning the ML models, and specifically the XGB algorithm, to determine the optimal hyperparameters, a *GridSearchCV* was employed with a 5-fold cross-validation strategy associated with the following parameters' ranges: *n_estimators*: [50, 100, 200], *max_depth*: [3, 5, 7], *learning_rate*: [0.01, 0.1, 0.2], *subsample:* [0.7, 0.8, 1.0], *colsample_bytree:* [0.7, 0.8, 1.0], *gamma:* [0, 0.1, 0.2], *alpha:* [0, 0.1, 0.5, 1], *lambda'*: [0, 0.1, 0.5, 1], resulting in 58,320 combinations subdivided into five cross-validation folds. It should be noted that *alpha* and *lambda* are the regularization terms, where *alpha* represents the L1 regularization term and *lambda* represents the L2 regularization term. In this regard, L1 regularization (Lasso) adds a proportional penalty to the absolute value of the coefficients, promoting sparsity in the model. L2 regularization (Ridge) adds a penalty proportional to the square of the coefficients, ensuring that all coefficients are small and preventing overfitting by avoiding extreme values. To prevent the features' magnitude from playing a role in the training process, the input features *(N*, *f*, *W)* were normalized using *Min–Max* Scaling to ensure that all features lie within the range [0,1]. Moreover, after the identification of the best hyperparameters for the XGB regressor, the model was further evaluated using 5-fold cross-validation, ensuring that the selected model generalizes well to unseen data and by using the MAPE as the evaluation metric. The XGB code was implemented in a Python/Spyder 3.11 environment considering *pandas*, *numpy*, *sklearn*, and *xgboost* libraries on the same workstation.

As concerns the DNN model, a feedforward architecture was employed. The network consists of an input layer with dimension 3, the same as the number of features (*N*, *f*, *W*), four hidden layers, and an output layer. Each hidden layer is based on the rectified linear unit (*ReLU*) in a 128, 64, 32, and 16 neurons configuration and with a dropout associated with a rate of 0.2. The output layer is defined as a single neuron with a linear activation function to generate continuous-valued predictions. The DNN model was trained with a learning rate of 0.0001, by employing the adaptive moment estimation (*Adam*) optimizer, and with a loss function based on the MAPE. A batch size of 32 was chosen to balance computational efficiency and training stability, and the network was trained for 30,000 epochs to ensure convergence. To validate the model's robustness, a 5-fold cross-validation strategy was adopted. As also carried out for the XGB model, features and target variables were normalized using *Min–Max* scaling, ensuring that all values ranged between [0,1], thereby preventing scale-related biases during training. Similarly, the MAPE was computed at each epoch, and for both training and validation sets for the $k=5$ cross-validation process, allowing for the reliable and robust training of the DNN model. The DNN code was implemented in a Python/Spyder 3.11 environment considering *pandas*, *numpy*, *sklearn*, and *keras* libraries on the same workstation.

Summarizing, the ME and ML models' parameter optimization times are equal to XGB: 4781 s, DNN: 659 s, DE: 4462 s, and CMA-ES: 1515 s, respectively, when run on a workstation with a 13th Gen Intel(R) i5-13500 (2.50 GHz) CPU and 64 GB of RAM. No parallel CPU-GPU computing was employed.

2.6 GMM test data generation

Considering the small size of the database employed in this research (Table 3), no experimental cases were removed beforehand and employed as test data. However, to provide an additional layer of validation while employing all the experimental results for the training of ME and ML models, synthetic data points were generated through the Gaussian mixture model (GMM) [45–47]. GMM synthetic data are distinct points generated within the experimental feature domain considering the Gaussian distributions of the experimental dataset (Table 3) and are used only for interpolation test purposes, not for extrapolation. Being independent points but belonging to the same features distribution, the GMM-generated data allow investigating the interpolation performance of ME-optimized regression functions and ML models within the feature latent space and without the need to create a separate test dataset. The GMM is a powerful technique that has been successfully employed in previous research by the authors for the generation of additional data points for intermediate validation in sheet metal forming [45] and database balancing in the presence of few independent cases and large number of features and target variables in biomechanical analysis scenarios [47]. Unlike bootstrapping or random resampling, the GMM approach preserves the multivariate correlations between features, ensuring that synthetic points realistically reflect the joint distribution of the process parameters within their original latent space [50, 51].

When employing GMM, the implicit assumption is that the original dataset, namely the independent cases within it, can be modeled by a combination, or mixture, of Gaussian distributions, each one characterized by a mean μ and a covariance Θ, as in Eq. (18).

$$p(x) = G(x|\mu, \Theta) \tag{18}$$

When all the z-distributions are considered together, the probability of x given z is assigned by the coefficient π_z and defined as in Eq. (19), based on marginal and conditional distributions as in Eq. (20).

$$p(x,z) = \sum_{z} p(z)p(x|z) = \sum_{z=1}^{Z} \pi_z G\left(x|\mu_z, \Theta\right), where \sum_{z=1}^{Z} \pi_z = 1 \tag{19}$$

$$Marginal \rightarrow p(z) = \int p(z,x)dx; Conditional \rightarrow p(x|z) = \frac{p(x,z)}{p(z)} \tag{20}$$

In this research, the total number of Gaussian distributions (Z) was fixed to $Z=5$, and the GMM model was trained considering the *log-likelihood* function through the expectation–maximization algorithm of Eq. (21).

$$\log L(X|\pi, \mu, \Theta) = \prod_{g=1}^{G} \log\left\{ \sum_{z=1}^{Z} \pi_z G\left(x_G|\mu_z, \Theta_z\right) \right\} \tag{21}$$

The ten GMM-generated points are displayed in Fig. 4, together with the raw data of Table 3, and summarized in Table 4. Considering the three features employed in this research, two plots are relevant to the experimental and synthetic data for TS and the other two for WR, respectively.

3 Results

3.1 FSP experimental results interpretation, statistical analysis, and modeling

3.1.1 FSP experimental results interpretation

Considering the complexity of the interaction between the N, f, and W features and TS and WR target variables, as reported in Table 3, average features dependency plots were generated for TS and WR, as shown in Fig. 5. In all six subplots, the independent variable is taken directly from the experimental values, while the dependent variable represents the mean response of the output (either TS or WR) at each level of the corresponding factor, averaged across all combinations of the remaining variables. In other words, taking for instance in Fig. 5a, the plotted point for $N=1200$ rpm corresponds to the mean TS measured in all experiments where $N=1200$ rpm considering all combinations of f and W at that level of N.

The interpretation of the results presented in Fig. 5 is grounded on the established process-structure–property relationships in the FSP process and the strong influence of process parameters and reinforcement on microstructure,

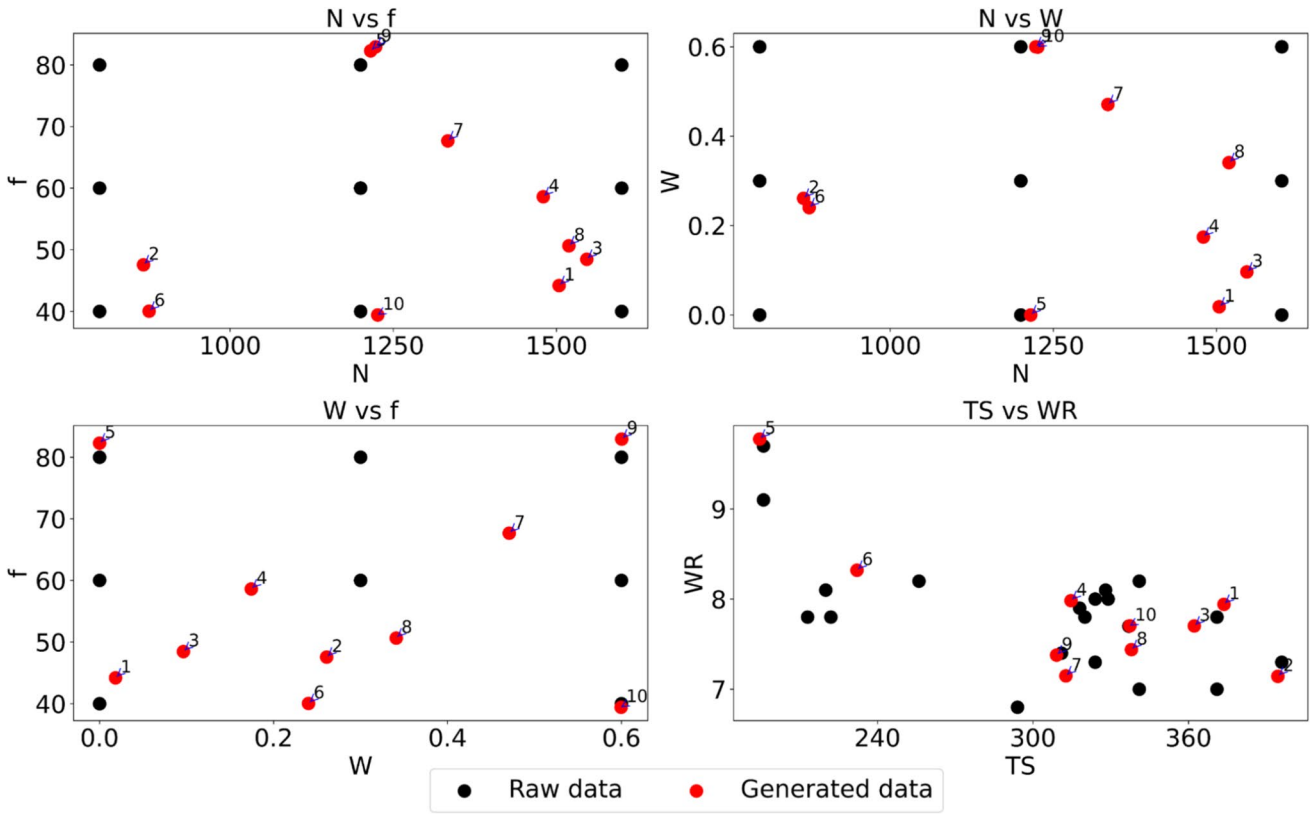


Fig. 4 Two-dimensional (2D) distributions for experimental data (Table 3) and GMM-generated synthetic data, the latter employed to test the prediction performance of the trained ML and ME models

Table 4 GMM-generated data points for the interpolation performance evaluation of ME-optimized regression functions and ML models

#	*N* (rpm)	*f* (mm/min)	*W* (g)	TS (MPa)	WR·10^{-3} (mm^3/m)
1	1504	44	0.0	374	7.9
2	867	48	0.3	395	7.1
3	1547	48	0.1	362	7.7
4	1480	59	0.2	315	8.0
5	1215	82	0.1	195	9.8
6	876	40	0.2	232	8.3
7	1334	68	0.5	313	7.1
8	1519	51	0.3	338	7.4
9	1223	83	0.6	309	7.4
10	1226	39	0.6	337	7.7

which in turn govern the observed mechanical (TS) and tribological (WR) properties [3, 19].

In Fig. 5a, the tensile strength (TS) follows a nonlinear trend with tool rotation speed (*N*), the latter governing the heat generation due to the friction interaction between the tool, the reinforcement particles, and the base material workpiece [19, 52]. Specifically, TS increases from 800 to 1200 rpm but begins to decrease at 1600 rpm, forming a quadratic behavior. This behavior is a classic example of process–structure–property relationship in the FSP of MMC where moderate *N* (1200 rpm) allows achieving sufficient plastic flow and dynamic recrystallization, resulting in finer grain size and uniform particles dispersion. On the other hand, at higher N (1600 rpm), excessive heat leads to coarser grain and Al_2O_3 particles' clustering or ejection from the stirring area, degrading both the reinforcement and its effectiveness, thus ultimately leading to a reduction of TS.

These results suggest that moderate rotation speeds optimize heat generation and plastic flow, while excessively high speeds lead to thermal softening and structural degradation [52]. Up to 1200 rpm, friction and plastic deformation are the main contributing factors, promoting uniform particle adhesion with the base material, thereby improving strength [53]. However, when the tool's rotational speed increases beyond 1200 rpm and reaches 1600 rpm, the centrifugal forces become predominant, leading to a decrease in the effectiveness of the reinforcement and in terms of the resulting tensile strength [14].

On the other hand, TS also shows a strong reduction for increase (*f*) (Fig. 5b), a fact caused by the reduction in the contact time between the tool and the workpiece, leading to insufficient plastic deformation and poorer Al_2O_3 dispersion, both negatively influencing factors for TS. In the FSP process, *f* controls the heat generation due to the friction interaction between the tool and the workpiece,

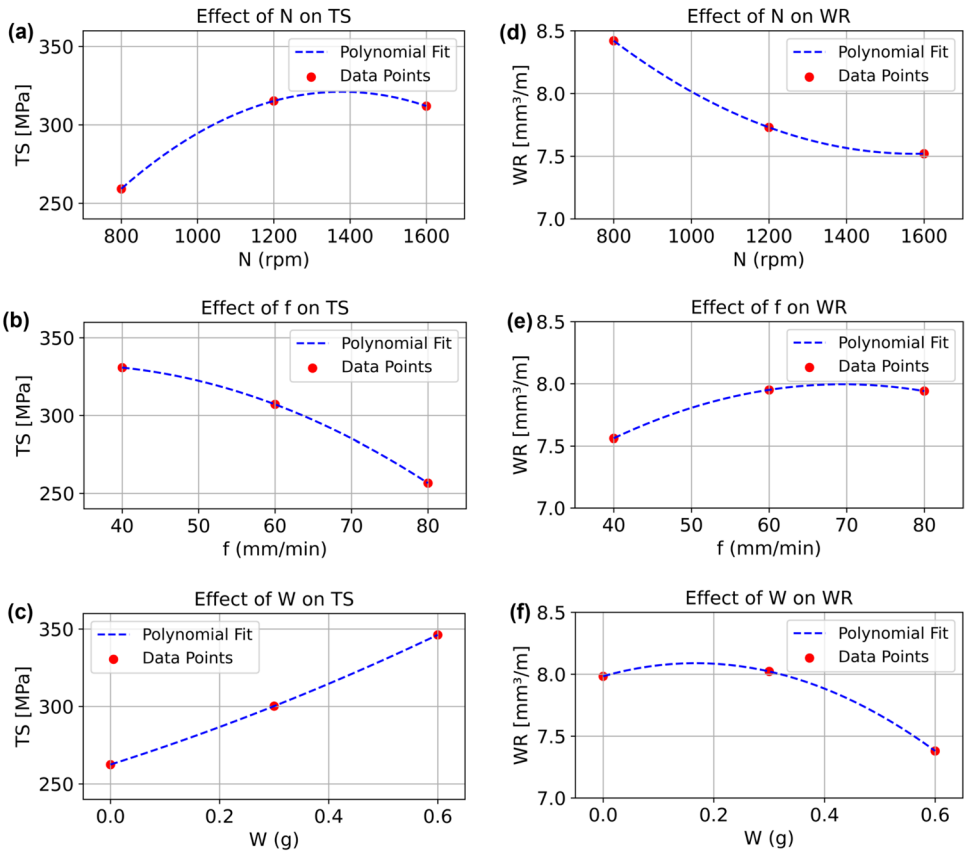


Fig. 5 Effect of the FSP process parameters, in terms of tool's rotational speed, feeding speed, and Al_2O_3 reinforcement weight on the tensile strength (TS) (**a**), (**b**), (**c**) and wear rate (WR), (**d**), (**e**) (**f**)

the reinforcement adhesion with the base material, and the process time [11]. A high feed rate reduces heat generation, with a consequent reduction in the plastic deformation and thus limiting the ability of the Al_2O_3 to adhere with the base material. Besides, from the microstructural point of view, a high *f* results in larger grain size with less uniformity throughout the whole workpiece, with the secondary effect of weak matrix-particles interface strength, thus in lower TS [18]. In such conditions, the dominant effect resembles more ploughing through the reinforcement rather than stirring, a fact that leads to the poor dispersal of Al_2O_3 particles in the stirred zone, reducing TS, as shown in Fig. 5b.

In contrast, TS increases almost linearly with the Al_2O_3 weight fraction (*W*) (Fig. 5c). A higher number of particles helps mitigate the negative effect of a high *f* on TS, allowing for more particles to adhere to the base material. In addition to that, a higher *W* also results in a finer and more uniform dispersion, creating more sites for the nucleation of dynamic recrystallization, helping in the refinement of the matrix grain size and providing more obstacles in the movement of dislocations [7, 39]. All these effects collectively help in improving the effectiveness of the Al_2O_3, ultimately improving TS, as shown in Fig. 5c.

Similarly to the consideration drawn for TS, the wear rate (WR) decreases with tool rotation speed (*N*) up to 1600 rpm (Fig. 5d), indicating that although a higher rotational speed for the tool increases the amount of ejected particles outside of the stirring zone, this issue does not affect the outer layers of the MMC, thus the decrease in WR, but mostly affects the depth of the particles penetration within the base material, hence the reduction in TS shown in Fig. 5a. Although excessive W may cause agglomeration or the clustering of particles, leading to degradation [8, 19], in the W range employed in the current study, the TS(W) correlation shows a monotonically growing trend, highlighting no issues of excessive particles agglomerations.

Considering now the results presented in Fig. 5b and e, higher feed rates result in a reduction of TS and an increase and plateau for WR. For higher f, lower heat generation, higher cooling, and the dominance of ploughing through the reinforcement rather than stirring result in a generalized reduction of the particles' effectiveness for both TS and WR [9, 16]. From the microstructural standpoint, higher *f* leads to larger and elongated grains and particles clustering, leaving exposed parts of the base materials, with a resulting lower wear resistance. In this regard, the negative effect of higher *f* is more pronounced on TS with a clear quadratic behavior from 40 to 80 mm/min (Fig. 5b), than on WR, the latter showing an asymptotic behavior from 60 mm/min.

Interestingly, WR exhibits a strong parabolic relationship with Al_2O_3 weight (W) (Fig. 5f) rather than a continuous decrease. The wear rate is almost unaffected by the addition of 0.3 g of Al_2O_3 particles, but doubling this amount results in a drop from approximately 8 to 7.4 mm^3/m, showing the positive effect of adding more reinforcement. In other words, the identification of the minimum reinforcement amount is critical to achieve a uniform protective layer on the top of the base material. In this regard, it ought to be pointed out that the drop in WR is higher for the case of higher values of *N* rather than higher *W* (Fig. 5d vs. f), meaning that the FSP process parameters have a higher influence in controlling the WR than W, as also shown in similar literature studies [4, 14]. On the other hand, for TS, the effect of W is higher than that of *N* (Fig. 5c vs. a), meaning that the amount of reinforcement has a stronger influence on the mechanical response [5].

Among the experimental conditions considered in this work, as reported in Table 3, the highest TS of 396 MPa and the lowest WR of 7.3×10^{-3} mm^3/m were achieved with $N = 1200$ rpm, $f = 40$ mm/min, and $W = 0.3$ g (sample #11). Both these results supersede the TS values for the AA1100 base material, TS = 197 MPa, as well as the unreinforced FSP, TS = 294 MPa. Moreover, these results are consistent with other literature studies where the optimal process conditions for the FSP are obtained at moderate tool rotation speeds and lower feed rates, conditions that maximize plastic flow and promote homogeneous particle distribution [14, 26]. Conversely, the worst combination of process parameters, leading to the lowest TS of 196 MPa and a WR of 9.7×10^{-3} mm^3/m, was observed for $N = 800$ rpm, $f = 80$ mm/min, and W = 0 g. In this scenario, the lack of reinforcement and the relatively high feed rate led to mechanical properties that are indistinguishable from the unprocessed base material, as similarly reported in the studies of Thangarasu et al. [13] and Devaraju et al. [14].

Overall, although scanning electron microscopy (SEM) analysis was not carried out in this research, the trends observed in Fig. 5 are fully consistent with the current knowledge related to the FSP. Tool rotation and feed rate mainly influence the thermal profile and the stirring efficiency, which in turn are related to grain size, particle distribution, and the matrix-reinforcement interface quality [4, 11]. In addition to that, although the amount of reinforcement plays an important role, minimum and maximum thresholds should not be exceeded to avoid either poor particle distribution or agglomeration, both resulting in poor MMC performance [3].

3.1.2 ANOVA analysis and TS-WR modeling

The ANOVA analysis was carried out in the Python/Spyder 3.11 environment on the experimental database of Table 3 together with the regression function form of Eq. (1) and by considering separately TS and WR and target variables.

The results are reported in Table 5 and hereafter commented upon.

For TS, the percentage contributions of tool rotation speed (N), feed rate (f), and Al_2O_3 weight (W) were found to be 19.07%, 1.29%, and 0.71%, respectively. This suggests that tool rotation speed (N) has the highest effect on TS, whereas the feed rate (f) and Al_2O_3 weight (W) contribute minimally. Additionally, the quadratic term of tool rotation speed (N^2) accounts for 16.51% of the variance, further confirming its strong nonlinear influence on TS.

This indicates that the relationship between TS and tool rotation speed (N) is more than linear with complex effects on TS. In addition, a high residual is associated with TS, suggesting that a significant portion of the variability is not captured by the selected regression model, summarized in Eq. (22), based on the polynomial form of Eq. (1). The second-order polynomial regression of Eq. (22) is associated with a root mean square (RMS) percentage error of 8.72% and an interpolation accuracy (R^2) of 74.7%.

Both high residual and relatively low interpolation performance, in terms of RMS and R^2, can be caused by a combination of factors, among which measurement uncertainties and the presence of more than quadratic effects, not included in (1). These considerations are aligned with other literature observations highlighting the need for more complex or higher-order models to grasp the complex features-target variable interactions [14]. Though the inclusion of higher-order terms might have a positive, yet fictitious, effect on regression accuracy, the optimization of the regression equation form, rather than its coefficients, is out of the scope of this research since the objective is to optimize the equation coefficients through ME modeling.

$$TS[\text{MPa}] = 268.23 + 229.449N - 136.568f + 13.829W - 172.649N^2 + \\ +29.318f^2 + 47.775W^2 - 12.5N \cdot f + 4.5N \cdot W + 70.606f \cdot W \quad (22)$$

As concerns the ANOVA results for WR (Table 5), the percentage contributions of tool rotation speed (N), feed rate (f), and Al_2O_3 weight (W) were found to be 18.55%, 5.82%, and 0.93%, respectively. This suggests that tool rotation speed (N) remains the dominant factor influencing WR, whereas feed rate (f) has a moderate effect, and Al_2O_3 weight (W) plays a minimal role. In addition, the influence of the quadratic terms is strong, especially for $N{\cdot}W$ and $f{\cdot}W$, calculated at 16.8% and 22.5%, respectively. In particular, the interaction between feed rate and Al_2O_3 weight ($f{\cdot}W$) suggests that the combined effect of feed rate and reinforcement content has a more significant influence than either factor individually. The empirical model for the prediction of WR is reported in Eq. (23) and is associated with an RMS percentage error equal to 1.65% and an interpolation accuracy (R^2) of 92.7%.

$$WR[\tfrac{mm^3}{m}10^{-3}] = 8.36 - 3.812N + 2.5749f + 0.802W + 2.413N^2 + \\ -1.185f^2 - 1.502W^2 - 0.5N \cdot f + 1.5N \cdot W - 1.678f \cdot W \quad (23)$$

The combined considerations of the ANOVA analysis for TS and WR indicate that N is the dominant factor in controlling the behavior of both target variables, and especially TS, as also suggested in previous studies [14, 26]. In a similar way, the effect of the Al_2O_3 weight (W) on TS is also consistent with the established literature of the FSP, where the effect of increasing W showed less influence than changes in the process parameters [13]. For WR, the significance of the interaction term ($f{\cdot}W$) suggests that maximizing wear resistance requires a combined optimization of both process parameters and reinforcement content to obtain a uniform distribution of hard particles capable of minimizing the surface wear through an effective protective layer [26].

Table 5 ANOVA results for TS and WR target variables

Feature	ANOVA results for TS				ANOVA results for WR			
	Adj SS	% Contr	F-value	P-value	Adj SS	% Contr	F-value	P-value
N	6479.9	19.1	3.5980	0.0871	1.2	18.6	19.3067	0.0013
f	439.4	1.3	0.2440	0.6320	0.4	5.8	6.0599	0.0336
W	243.0	0.7	0.1349	0.7211	0.1	0.9	0.9671	0.3486
N^2	5610.0	16.5	3.1150	0.1000	1.1	16.3	16.9966	0.0021
F^2	109.9	0.3	0.0610	0.8098	0.2	2.7	22.7912	0.1257
W^2	335.0	1.0	0.1860	0.6754	0.3	4.9	5.1380	0.0468
$N{\cdot}f$	78.1	0.2	0.0434	0.8392	0.1	1.9	1.9398	0.1939
$N{\cdot}W$	10.1	0.0	0.0056	0.9417	1.1	16.8	17.4578	0.0019
$f{\cdot}W$	2667.8	7.9	1.4813	0.2515	1.5	22.5	23.3994	0.0007
Residual	18009.8	53.0	-	-	0.6	9.6	-	-

As a final remark, the results of the preliminary investigations relevant to the AIC (Akaike information criterion) and BIC (Bayesian information criterion) analyses from the first- to fourth-order polynomial regressions for both TS and WR, as well as the percentage errors for second- and third-order polynomial regressions for the 20 points in Table 3, are reported in the Appendix of the paper.

3.2 Regression coefficients optimization by ME models and TS-WR modeling

Considering the regression function presented in Section 3.1, this section aims to provide the results of the application of the ME models in the optimization of the θ coefficients for Eqs. (22) and (23) and the relevant deviations with respect to the true values reported in Table 3. The results, summarized in Fig. 6, offer a clear and exhaustive overview of *pro et contra* related to the ME-based optimization, as well as strengths and weaknesses of the ANOVA-based second-order regression functions of (22) and (23).

First, in Fig. 6, the error spikes for three instances were left out of the plot, and the relevant percentage errors are reported separately to provide a more focused view of the majority of the results. In addition, the average percentage errors are reported at the top of the relevant bars on the far right of both charts to provide a more direct comparison of the three models' performance. Considering both charts together, the difference in magnitude for the percentage errors related to TS, to WR, is related to the difference in residuals as shown from the ANOVA analysis of Section 3.1, specifically in Table 5. In this regard, WR is well represented by Eq. (1) with a residual lower than 10%, while TS has a residual higher than 50%, meaning that part of the relationship between this target variable and the three selected features is most likely of an order higher than the second. In addition to that, concerning TS, both the base regression and the ME-optimized models show high deviations for a limited number of cases, while generally the latter have lower deviations overall.

Considering the results in Fig. 6a, for the DE-optimized model, apart from two error spikes with errors greater than 30%, and one at approximately 22%, the remaining deviations are well below 10%, with an average of 5.6%. In contrast, the average deviations for the CMA-ES model are 28.5% higher, with an average error of 7.2%, while those for the basic regression function are 55.3% higher and associated with an average percentage error of 8.7%. In Fig. 6a, the high deviations are clustered between cases #9 and #13, which include the minimum and maximum values for TS, namely #9 at 196 MPa and #11 at 396 MPa, respectively. In this regard, the DE-optimized regression function predicts remarkably better than the base regression and CMA-ES-optimized model for both cases #9 and #11 alike, a fact which is also confirmed for the case of WR (Fig. 6b).

As concerns the average percentage errors for WR, a similar trend as that highlighted for TS is identified, with the regression function based on the DE-optimized θ coefficients having the best performance. It is worth noting that for WR all three sets of coefficients lead to one case with a high percentage error of around 8%, but the lower deviations granted by the DE-optimized coefficients allow for the average error to be limited to 1.1%, similar to the 1.2% for the

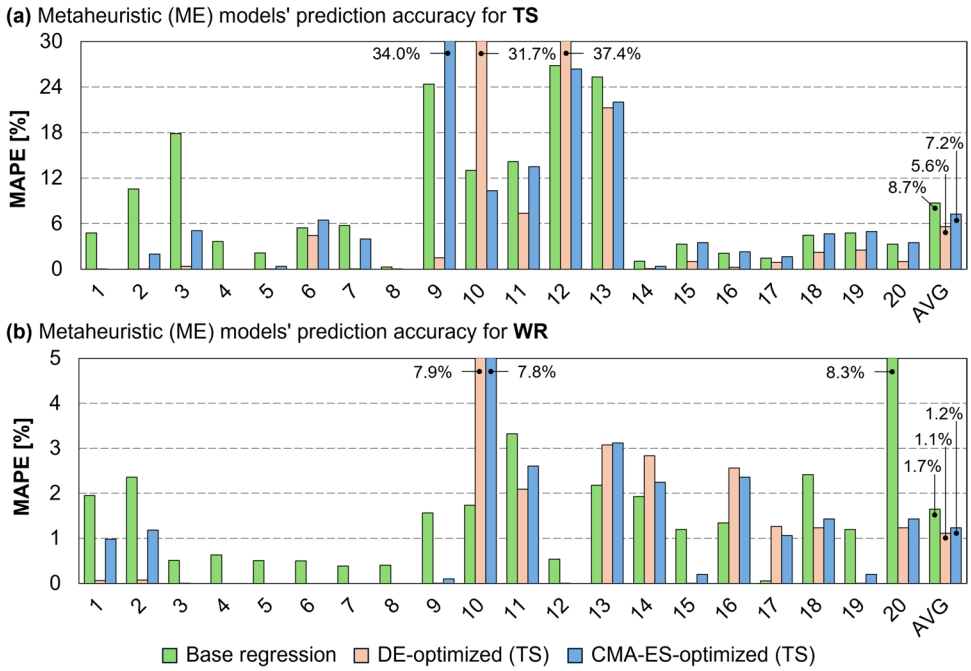


Fig. 6 Comparison among base regression, DE-optimized, and CMA-ES-optimized coefficients for **a** TS and **b** WR

CMA-ES, but 54.5% lower than that of the base regression coefficients, and equal to 1.7%.

From a global perspective, as concerns the ME-optimized coefficients for the second-order regression function of Eqs. (22) and (23), the DE-optimized regression function showed to be consistent for both TS and WR, with the lowest average percentage errors as well as mostly stable and lower deviations across all 20 cases. Besides, the performance of the CMA-ES-optimized coefficients shows a strong improvement to the coefficients resulting from the ANOVA analysis as well, confirming the effectiveness of the applied methodology, regardless of the considered formulation.

The results presented in this section align well with recent literature on data-driven process optimization where ME approaches, though not yet extensively employed for FSP modeling, are increasingly recognized for their capacity to improve parameter selection and surrogate model fidelity, especially when dealing with nonlinear and complex structure–property relationships. As reviewed by Rajwar et al. [54], DE and CMA-ES algorithms have been successfully applied in a wide range of engineering optimization problems due to their ability to avoid local *minima* and identify global solutions, as well as enhance the performance of surrogate models and supervised learning frameworks [55, 56]. While direct applications of DE or CMA-ES to FSP parameter optimization remain an open research area, the positive effect shown in optimizing the coefficients of Eqs. (22) and (23) is consistent with literature findings showing the capability of both formulations in avoiding local *minima* and providing robust solutions also for small dataset scenarios.

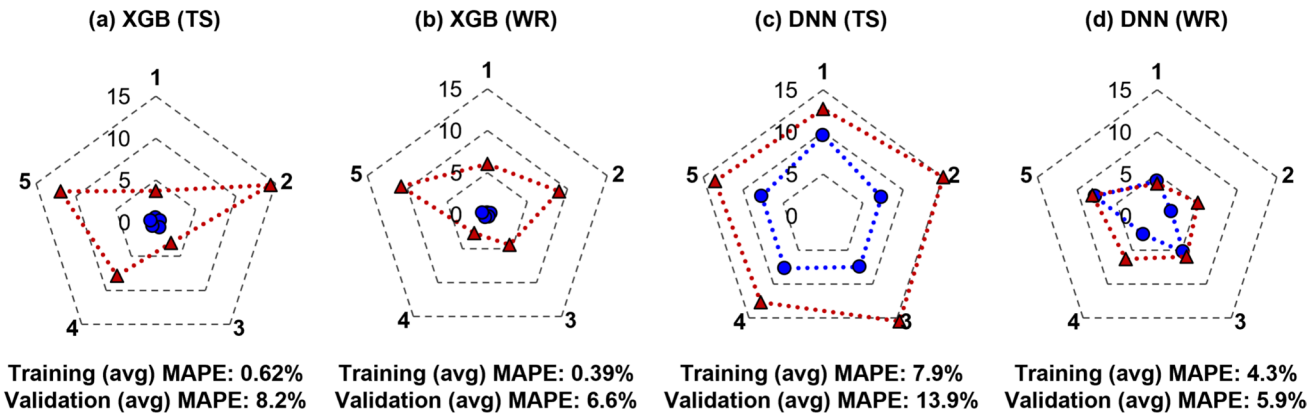
3.3 TS-WR modeling by XGB and DNN ML models

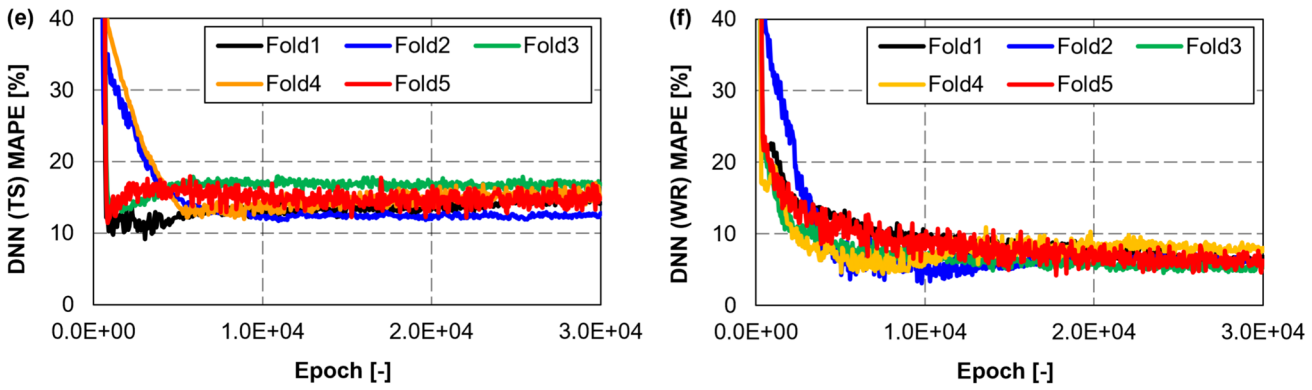
The application of the supervised XGB and DNN ML models, as presented in this section, provides an alternative approach to the ME-optimized regressions discussed in Section 3.2. The optimized hyperparameters for the XGB, as resulting from the *GridSearch* analysis presented in Section 2.5, are reported in this section, and having considered a single-variable regression model, different hyperparameters are obtained for TS and WR models. For TS, the best hyperparameters are *n_estimators*: 100, *max_depth*: 3, *learning_rate*: 0.2, *subsample*: 1.0, *colsample_bytree:* 0.7, *gamma:* 0, *alpha:* 1.0, and *lambda'*: 0.1, while for WR: *n_estimators*: 50, *max_depth*: 3, *learning_rate*: 0.2, *subsample*: 1.0, *colsample_bytree:* 1.0, *gamma:* 0, *alpha:* 0, and *lambda'*: 0, respectively. The detailed MAPE results relevant to the training and $k=5$ cross-validation are reported in Fig. 7a–d.

From a general perspective, both XGB (Fig. 7a, b) and DNN (Fig. 7c, d) models show good training and fairly good validation performance, with the former exhibiting particularly low training error but higher variance across cross-validation folds, close to 15%. In this regard, high variability in the validation deviations and consistently higher magnitude in comparison to the training score highlight an overfitting risk scenario, a well-known challenge in *tree-based* ensemble methods when applied to small datasets [32, 36, 44]. Although the XGB model has a better tradeoff between variance and bias than other ensemble algorithms [44], it is still susceptible to overfitting in the case of small-scale databases, as the one employed in this research. In contrast, the DNN model produced more balanced and consistent errors across folds, particularly for WR, reflecting the robustness of deep learning architectures [33]. This behavior is consistent with literature results showing the better adaptability of the DNN to ensemble methods, such as XGB, when it comes to complex and nonlinear relationships between input features and target variables, especially for the case of small datasets [33, 45]. Nevertheless, DNN accuracy is still subordinate to proper regularization and prevention of divergent behaviors during the training, a fact which must be either prevented by early stop or verified by MAPE of MAE vs. epoch trends analysis. In summary, considering the training and validation MAPE across all five folds (Fig. 7a to d), the XGB model shows better performance for TS, though the different deviations across the validation folds suggest overfitting, whereas the DNN is more balanced for both TS and WR, with better performance than the XGB model for the latter target variable.

Considering the results presented in Fig. 7a–d, and in light of the small size of the database employed in this research, the good agreement among train and validation MAPE, for both TS and WR modeling, allowed concluding that no outliers are present among the 20 experimental cases reported in Table 3. This is consistent with similar literature scenarios where the careful designing of the experimental cases allowed minimizing, or avoiding completely, the presence of outliers in the training dataset [57]. Moreover, concerning the DNN model, the MAPE at the end of the training process, concluding at 30,000 epochs, sees fairly uniform values for both TS and WR, although the latter is more precise (Fig. 7e and f). This fact is well correlated with the training and validation MAPE reported in Fig. 7c and d, which shows that the DNN model for WR is consistently more accurate than that of TS, though both stably converge to similar error magnitudes at the end of the learning process.

In addition, a similar comparison to that carried out for the ME models is proposed in this section. To this end, the results presented in Fig. 8 are related to the best out of the 5 models resulting from the *k*-fold cross-validation process, applied for the prediction of all 20 cases, and relevant to both XGB and DNN. This approach ensures a fair benchmark with the ME models, which were both trained and validated on the full experimental set. Such a comparison is critical to isolate the effect of modeling strategy under the same data constraints, a known challenge in small-scale manufacturing datasets [45].





MAPE [%] @ 30,000 epochs

Fold1	Fold2	Fold3	Fold4	Fold4
14.1	12.8	16.9	15.1	14.9

MAPE [%] @ 30,000 epochs

Fold1	Fold2	Fold3	Fold4	Fold4
6.8	5.7	5.3	7.9	6.2

Fig. 7 Training and validation for the 5-fold cross-validation for **a** XGB (TS), **b** XGB (WR), **c** DNN (TS), and **d** DNN (WR) machine learning models. MAPE vs. epoch for **e** TS prediction and **f** WR prediction for the 5-fold cross-validation process of the DNN model

First, it is evident that the magnitudes of the deviations for the ML models are globally lower than those of the ME-optimized regression functions counterparts, as visible by comparing Fig. 6 with Fig. 8. Besides, also for the ML models, the maximum MAPE for TS modeling is still localized between cases #7 and #11, similarly to Fig. 6a, but is approximately half the magnitude, ~ 20.9% for the DNN model in case #11, which also represents the maximum percentage error for TS. This fact indicates that the DNN model can interpolate more robustly even in regions of high output variability, as demonstrated in the literature where neural models have shown enhanced performance for nonlinear and multivariate process-property relationships [33, 45].

The deviations for the XGB model are similar to those of the DNN, but with an average value of approximately half the magnitude, 6.2% for DNN and 2.9% for XGB. Although counterintuitive, the slight overfitting issue mentioned before for the XGB model, visible from the difference between train and validation test scores in Fig. 7a and b, plays a pivotal role in reducing the deviations on the training dataset points, thus resulting in a lower average MAPE than DNN, but at the expense of generalization [44, 57], as it will be proven in Sect. 3.4. In addition to the overfitting issue, the XGB model still generates error spikes (nearly 29%) for WR in case #7, revealing a degree of instability for poorly represented and under-constrained regions of the latent space [31, 33].

From the analysis of the performance on the training dataset points between ME-optimized regression functions reported in Section 3.2 and ML models, the latter presented in this section, a few considerations can be drawn. First, from a general overview, the ML formulations seem to have lower average MAPE and also less tendency to sudden error spikes, which are more common for ME-optimized regression functions.

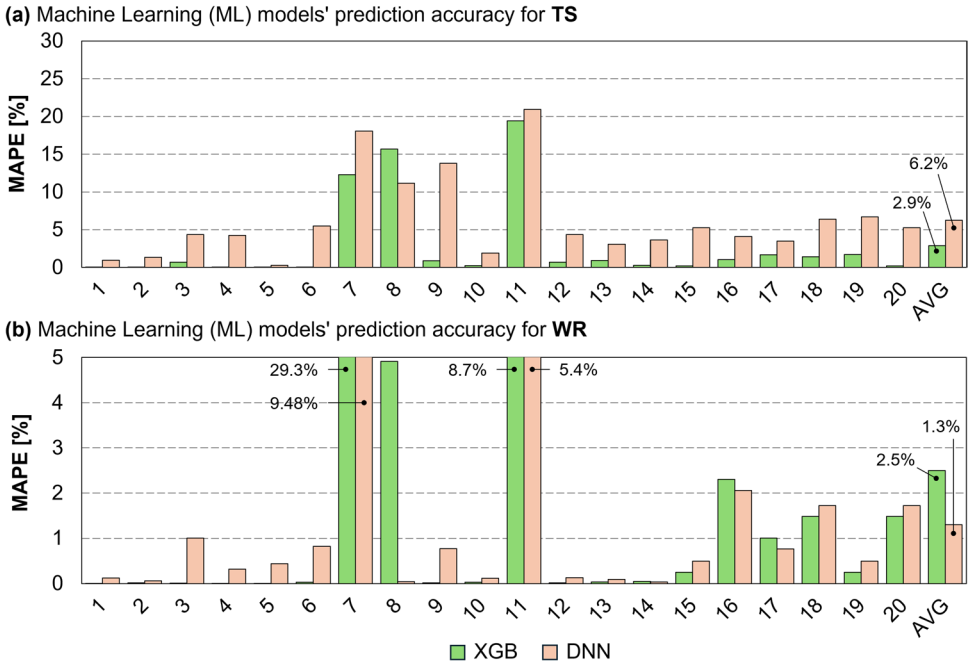


Fig. 8 Comparison between XGB and DNN models for the prediction of **a** TS and **b** WR

In this regard, and concerning the TS modeling (Fig. 6a vs. Figure 8a), this superiority is most likely caused by the high residual of the base regression function for TS (Table 5), which inherently limits the potential improvement achievable by ME methods through coefficient optimization alone. Although the consideration of a third-order base regression model would have resulted in better performance, it would have also resulted in overfitting on the training data point and almost negligible effect of the ME optimization, hindering even more generalization and the interpolation performance within the features space. Conversely, the differences between ME-optimized regressions and ML models are less pronounced for the WR modeling (Fig. 6b vs. Figure 8b), and both approaches show error spikes in particular cases, reflecting the intrinsic difficulties of modeling tribological outcomes in the presence of multifactorial effects [26, 33] and at the fringes of the features space.

Overall, the results presented in Sections 3.2 and 3.3 are aligned with the consensus that ML solutions show better adaptability in case of small-scale databases or complex manufacturing engineering datasets [33] and higher sensitivity of ME models to problem dimensionality [54]. Nevertheless, both classes of models are still susceptible to local error spikes due to data sparsity resulting in insufficient constraints during the training process.

3.4 ME-optimized and ML performance on GMM-generated cases

As summarized so far, both ME and ML models showed points of strength as well as limitations when tested on the 20 experimental cases employed for their training. However, the issue of prediction capabilities on new data is crucial in every engineering environment where the trained model is employed on an unseen combination of features to predict the relevant target variables. To this end, a GMM-generated synthetic dataset was defined, as described in Section 2.6, and employed to test both ME-optimized regression functions and ML models on unseen data, though belonging to the same distribution of probability of the experimental data.

To this end, the 20 experimental cases and the 10 GMM synthetic test data points are reported in Fig. 9 and Fig. 10, together with the contour plot of the MAPE error, with all minimum values set to zero and the maximum values according to each case. The results for the ME-optimized regressions and ML models for TS are reported in Fig. 9, whereas those for WR are in Fig. 10, respectively, and refer only to f and N features, both applicable in the presence or absence of reinforcement.

First, in light of the results of Fig. 9, and comparing the maximum MAPE as set in the four scales on the right of the contour plots, the DNN model allows for the lowest MAPE, followed by the CMA-ES-optimized regression function. The DE-optimized results (Fig. 9c) show almost a two times higher maximum MAPE for high feeding speed values and in the proximity of both experimental and synthetic data. Although both DNN and CMA-ES-optimized formulations showed higher errors when compared to the other modeling alternatives in their respective categories, as shown in Sections 3.2 and 3.3, when applied to new data, they show better prediction performance, meaning that their training is less overfitted on the data themselves and more capable of representing the features'

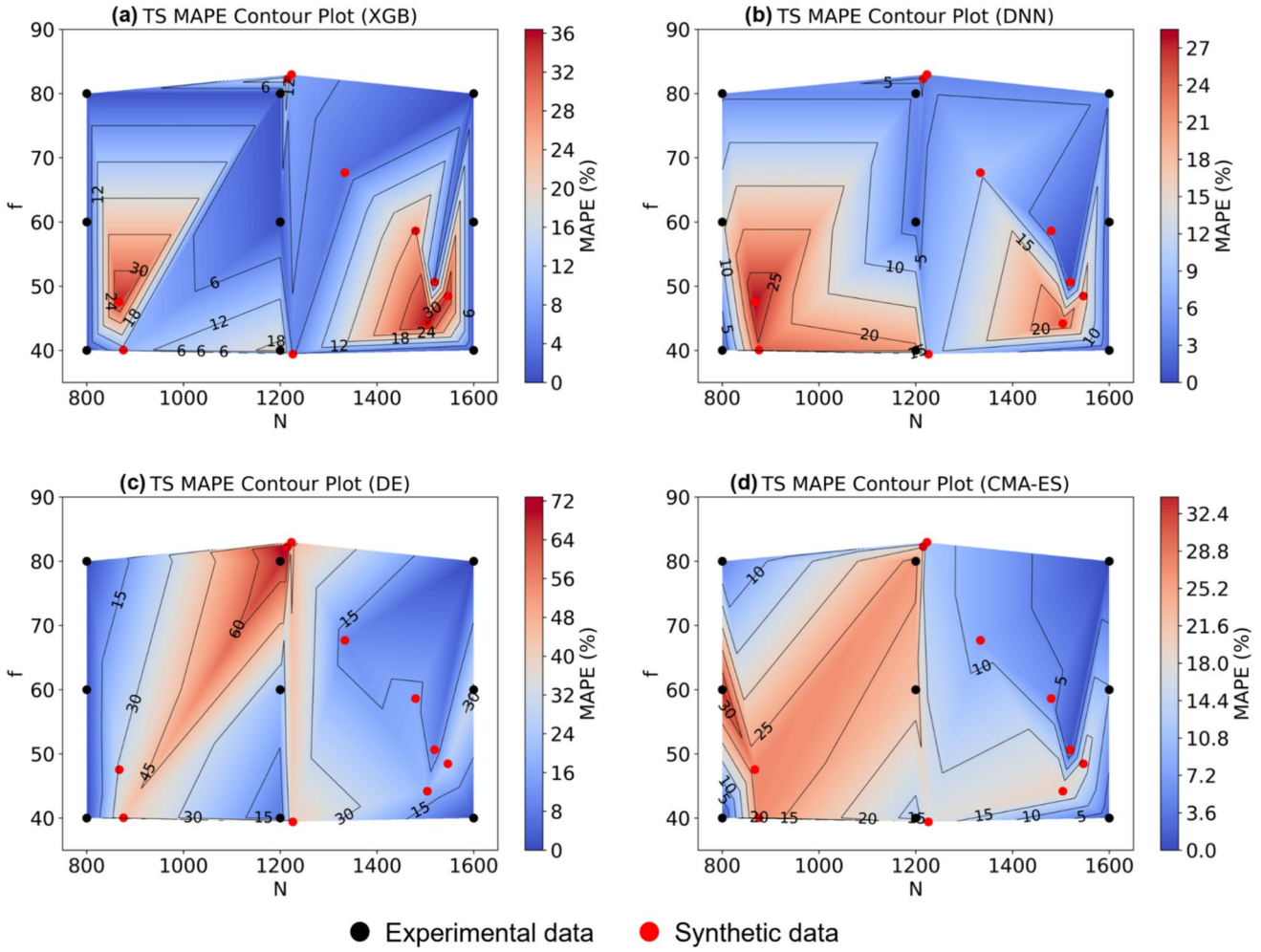


Fig. 9 Contour plots for the MAPE errors for experimental and GMM synthetic data for **a** XGB (TS), **b** DNN (TS), **c** DE (TS), and **d** CMA-ES (TS)

distribution and interpolating effectively within the latent space. These results are in good agreement with the consensus reported in recent literature showing that evolutionary strategies, such as CMA-ES, and neural networks, such as DNN, generalize better within the interpolative domain when trained on small or imbalanced datasets [33, 54]. Specifically, for DNN, regularizations help in modeling non-linear features interactions, typical of FSP and manufacturing engineering in general [33, 45]. As concerns CMA-ES, its population-based search is less likely to get trapped in narrow, local *optima* defined by the available points, thus promoting more robust solutions over unseen, yet plausible, input combinations [54, 55].

In addition to that, by comparing Fig. 9a with b, as well as Fig. 9c with d, the areas of the latent space where the MAPE distributions are remarkably similar, although with different magnitudes. This fact suggests that although the modeling approach differs, machine learning models tend to have limitations in similar sub-regions of the latent space, and the same applies also to the ME-optimization of a base regression function. In this regard, both XGB and DNN models have a fairly low MAPE on both experimental and synthetic data for high values of the feeding speed (f) and across the whole rotational speed (N) range, while they struggle for low values of f combined with high values of N. These limitations are inherent to the problem at hand, and specifically to the size of the employed database, rather than being related to specific features of the ME or ML models and have been observed also in other manufacturing process modeling dealing with scarce datasets [45, 56].

In terms of ME-optimized regression functions, the CMA-ES-optimized solution (Fig. 9d) shows good predictions within the features' boundary but struggles on its boundaries, a fact which is shared by almost all modeling techniques, ME and ML included. It is interesting to point out that the DE-optimized regression function (Fig. 9c) shows similar issues on the top boundary of the features space, though the magnitude is more than two times that of

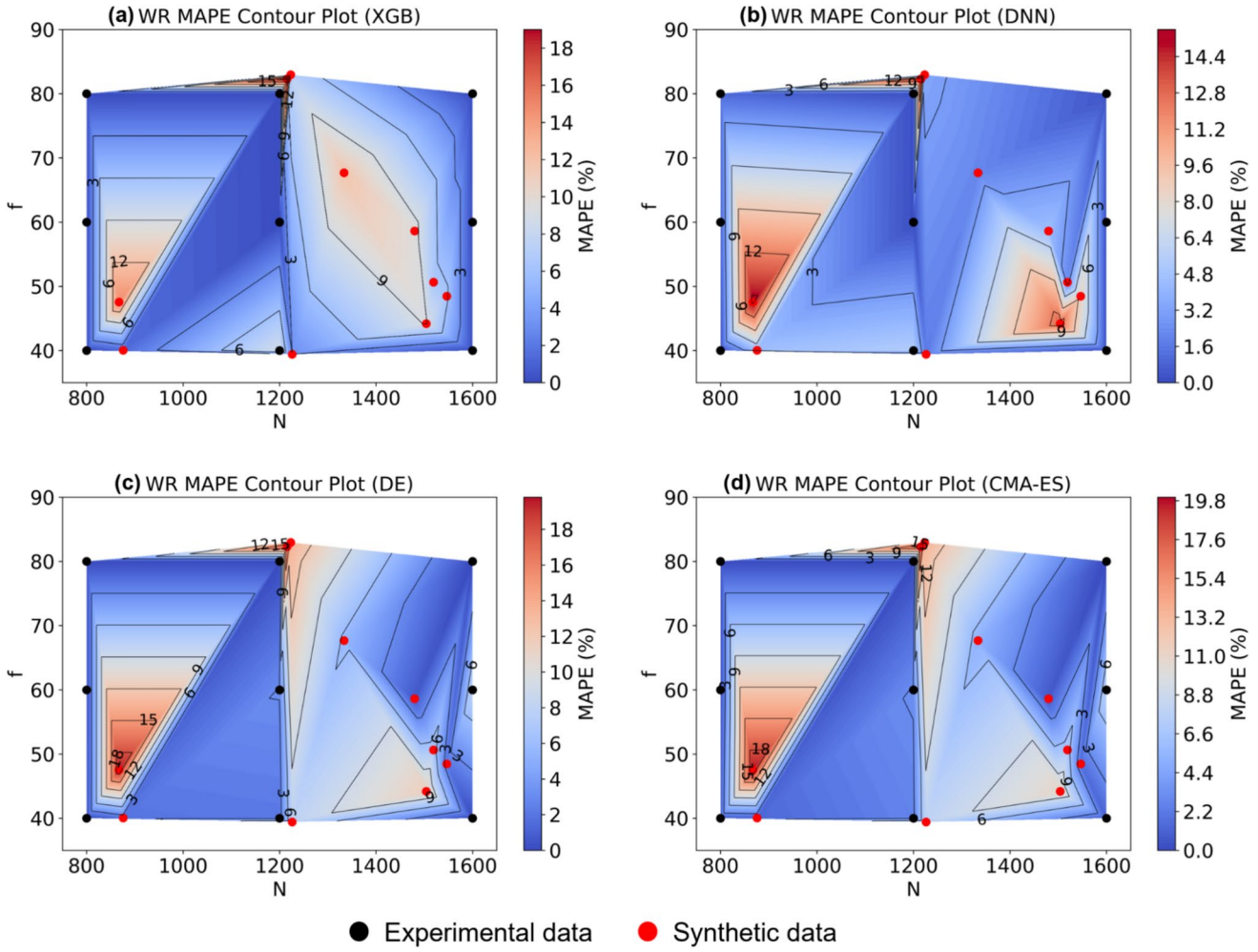


Fig. 10 Contour plots for the MAPE errors for experimental and GMM synthetic data for **a** XGB (WR), **b** DNN (WR), **c** DE (WR), and **d** CMA-ES (WR)

Fig. 9d, suggesting lower adaptability towards the fringes of the dataset, thus also implying limitations in extrapolation tasks. As concerns WR modeling (Fig. 10), the differences in MAPE magnitude are more subtle but still show the DNN model to have the lowest maximum deviation. In this regard, the high deviations for both TS and WR for the DNN model are located in similar sub-regions of the features space, specifically for low *f* and at the extremes of the *N* range, meaning that regardless of the target variable, the model is stable and robust across different target variables for the same features' distribution.

Globally speaking, when the ME-optimized solutions and the ML models are compared with one another only on experimental data, for the former, the DE solution regression function showed better performance than the CMA-ES ones. Similarly, the XGB model also resulted in lower deviations than the DNN, especially for the TS target variable. However, when new data are considered, though belonging to the same features' distribution of probability, the instability of those two solutions is evident from localized error spikes, especially at the domain boundary. On the other hand, the DNN model allows for better interpolation performance on unseen data within the features' domain, though unavoidable limitations persist at its boundary, with an expected further increase in the MAPE if extrapolation tasks were to be attempted.

In summary, the results presented in this section, together with those in Sections 3.2 and 3.3, suggest that, while ME-optimized regressions and *tree-based* ML (XGB) models may show high performance on training data, they have a limited generalizability within the FSP features space, especially in low data points regions [55, 56]. Conversely, DNNs, and to some extent CMA-ES, are more capable of better mapping the interpolation space between experimental data points, although none of the tested approaches is immune to error spikes at or beyond the convex hull of the training set [33, 54, 57].

4 Discussion

The results presented in the previous chapter offered an overview of the differences, and similarities, between ME-optimized polynomial regressions and ML model solutions' on both the experimental FSP database points as well as on the GMM-generated synthetic data. To provide a better insight into the relevant implications, this section aims to discuss further the results, focusing on the modeling techniques themselves and their integration with the FSP process. In addition to that, the limitations of the current study and possible future improvements are presented as well to provide a comprehensive summary and expected outlooks for this research.

The first critical aspect, also mentioned in Section 3, is represented by the need for an equation discovery phase for ME-formulations. Indeed, as shown in Fig. 6, the optimization of the θ coefficients resulted in an improvement in the MAPE, but error spikes beyond 30% are present across all formulations for TS. This fact highlights the most critical downside of polynomial equations, regardless of whether they are directly derived from the ANOVA or further optimized. Though the form of Eq. (1) is well-suited to model WR, for the case of TS, the high residual (Table 5) demonstrates that a second-order polynomial function is not enough to model the relationship between features and the TS target variable. A higher polynomial function might be helpful in limiting the residual, but, at the same time, the inclusion of a higher number of terms will inevitably affect the ME-optimization, as well as increase the risk of including low-importance terms. A high number of terms in a regression function, though relatable with a fictitious high correlation and low MAPE, might also be indicative of overfitting on the training points, with a consequent low interpolation capability, and even lower extrapolation capabilities [58]. This phenomenon has been widely reported in literature contributions analyzing regression modeling for manufacturing processes, where over-parameterization can lead to deceptively high accuracy (low MAPE, high R^2) on the data points with poor interpolation performance within the features space [33, 54, 58].

As concerns the choice of the base polynomial regression function of Eq. (1), it should be noted that the primary objective of this study was not to achieve a perfect fit on the 20 experimental points, but to provide a consistent baseline for benchmarking ME and ML models in interpolation tasks. Allowing the polynomial model to fit perfectly would undermine this goal and introduce bias into the model comparison. *Id est*, comparing third-order polynomial functions for TS and WR with the results of XGB and DNN models, would create an unfair advantage toward the former, also considering the need for cross-validation of the latter. For all these reasons, a second-order polynomial function was employed for the definition of the base regression models for TS and WR. Although not strictly related to the research question and results presented in this manuscript, the second- and third-order polynomial fitting for TS and WR, together with the relevant R^2 values, are reported in the Appendix.

The issue related to the choice between second- and third-order polynomial fitting discussed above is remarkably similar in its effect, though different in nature, to the overfitting issue relevant to the XGB model, as grasped from the training and cross-validation results presented in Fig. 7. As demonstrated in previous research [44], the XGB is a powerful supervised ML model, which can be adapted to single and multivariate regression tasks but is also highly sensitive to the size of the dataset and its information density. In the FSP database employed in this research, the low number of independent cases and the spacing between them make it complicated for an ensemble (*tree-based*) formulation, such as the XGB model, to predict on points at the fringes of the latent space as well as to interpolate within it. This is shown in Fig. 9a, where the prediction on the GMM synthetic points, for low f values, represents the regions of the bi-dimensional (2D) features' space where the XGB model has the highest deviations. On the other hand, although the DNN model shows similar weaknesses in similar regions of the features' space, the transition of the MAPE is smoother, showing better prediction capabilities in between experimental points and toward test points. This observation aligns with recent studies in advanced manufacturing, which report that *tree-based* ensemble methods, while powerful for high-dimensional problems, often struggle to generalize well in regions with sparse data [33, 44]. In a similar way, but for a completely different modeling technique, the DE-optimized regression function of Fig. 9c has abrupt changes in the MAPE error, which appear smoother for the CMA-ES-optimized function, as in Fig. 9d.

Concerning the ME optimization models, both DE and CMA-ES ME models were selected among a total of six formulations including the Big Bang-Big Crunch (BBBC), Crow Search Algorithm (CSA), Ray Optimization Algorithm (ROA), and Particle Swarm Optimization (PSO) models. Though some improvements in comparison to the baseline regression functions were observed, their performances were between one and two orders of magnitude lower than those granted by DE and CMA-ES formulations and, for this reason, were not included in this research. The superior performances of DE and CMA-ES observed here are consistent with recent literature in process optimization and surrogate modeling, which highlights the capability of these algorithms to effectively search complex solution spaces and avoid local *minima* even when experimental data are scarce or noisy [54, 55].

As concerns the ML formulations, though advanced and highly customizable, they still represent data-driven

modeling techniques, with no direct inclusion of any physics or prior knowledge related to the FSP process. Nevertheless, the DNN formulation showed to be capable of having reasonably low computational time associated with robust training and cross-validation scores, as testified by the MAPE across the five folds (Fig. 7) and on both experimental and synthetic data (Fig. 9 and Fig. 10). In addition to that, the DNN formulation showed relatively low MAPE and a smoother transition between experimental and synthetic data points. These facts, combined with the lowest training time of 659 s, suggest that if a purely data-driven approach is selected for the modeling of the FSP process, the DNN algorithm offers the best compromise, especially in interpolation tasks within the training features' domain on unseen data. This finding agrees with the recent literature showing that properly regularized and cross-validated DNNs can achieve high predictive accuracy in the modeling of complex manufacturing processes and under data-limited conditions [33, 45]. As a final remark, also for the case of the ML models, other formulations, other than the one presented here, were tested but deemed not accurate enough for the modeling scenario at hand and the small dataset considered. In this regard, the tested models included Random Forest, Gradient Boosting, Support Vector Machine regressor, and Multi-Layer Perceptron.

In addition to the consideration related to ME and ML models, as presented so far, the limited number of experimental data points (20 cases) employed in this research may represent a limitation in the generalizability of the results presented in this work. On the other hand, since both ME and ML modeling approaches were tested under the same data scarcity conditions, their performance in such an extreme scenario could be compared fairly and effectively. On top of that, incomplete and scarce data scenarios reflect a typical constraint in manufacturing engineering, where cost, time, and resource constraints often limit dataset sizes, as discussed in [31, 33].

In this research, and regarding the size of the employed database, the issue of data scarcity was addressed by generating GMM synthetic data belonging to the same combination of Gaussian distributions as the experimental data (Table 3). Although GMM-generated samples have proven to be reliable and effective in representing the overall features' distribution within the latent space, they are not real independent cases; thus, they cannot be employed to enhance the information density of the training dataset, but only for test purposes. This intrinsic characteristic of synthetic data is widely recognized in the literature and shows that such data are effective for benchmarking and stress-testing model generalizability but cannot fully substitute for independent experimental validation [50, 51]. Indeed, the consideration of more experimental points for training, validation, and especially test purposes would ultimately reduce even further the deviations of both ME-optimized regressions and ML models, although the general trend of the results would remain almost unaltered. On the other hand, such an approach is often impractical within the context of manufacturing engineering due to resource constraints and might be even inefficient if the training dataset must be prepared ad hoc [31].

In light of the results and considerations summarized so far, and considering the inferred complexity of the relationship between TS and the considered features, it is expected that the inclusion of physics-based governing equations or user-defined knowledge into the training scheme, in terms of Physics-Informed Neural Network (PINN) [33] or Gradient Enhanced-Expert Informed Neural Network (GE-EINN) [45], would have a positive effect on the prediction performance of the ML formulations. Although the utilization of more complex formulations, such as PINN and GE-EINN models, requires an equation discovery phase, the improvements in terms of interpolation and extrapolation performance have been already demonstrated for other manufacturing processes [33, 45], and it is expected to be advantageous also for FSP and FSVP modeling as well. On top of that, PINN or GE-EINN approaches can be combined with mitigation strategies such as bootstrapping, cross-validation, and synthetic data augmentation, as also partially carried out in this research, with the aim of identifying the best modeling technique for the investigated manufacturing process and considering the available training dataset.

Summarizing, although further research is indeed necessary to provide better guidance in the presence of limited raw data scenarios and complex interactions between features and target variables, the modeling techniques, results, and considerations presented in this work, and relevant for the FSP process of AA1100 base material reinforced with Al_2O_3 particles, provide a framework to develop reasonably accurate yet affordable models adaptable to a large variety of manufacturing engineering scenarios.

5 Conclusions

This study systematically benchmarked ME-optimized polynomial regressions and supervised ML models for predicting tensile strength (TS) and wear resistance (WR) in friction stir processed AA1100-Al_2O_3 MMCs, using a small-scale experimental dataset and synthetic GMM-generated data points for test purposes. The key findings are as follows:

- Optimal FSP parameter combinations ($N = 1200$ rpm, $f = 40$ mm/min, $W = 0.3$ g) can achieve TS increases of

20.7% and minimize WR to 7.3×10^{-3} mm^3/m, confirming previous literature trends.

- On experimental data, TS predictions see the ME-optimized regression functions show an average MAPE of 5.6% and 7.2% for DE- and CMA-ES-optimized regression functions, also associated with error spikes at 37.4% and 34.0%, respectively.
- Same concerning experimental datapoints, the ML formulation grants a reduction in both average MAPE, to 2.9% for the XGB and 6.2% for the DNN, also associated with lower error spikes, equal to 19.4% and 20.9%, respectively.
- Similar trends are also observed between the ME-optimized regression functions and ML models for the WR target variable. In this regard, average MAPE for DE- and CMA-ES-optimized regressions is equal to 1.1% and 1.2%, while for XGB and DNN, they are equal to 2.5% and 1.3%, respectively.
- However, on GMM-generated data, the DE-optimized regression and the XGB model show limited generalizability within the features domain due to a combination of overfitting issues and high residuals for the base regression model, based on a second-order polynomial function.
- On the other hand, CMA-ES-optimized regression and DNN ML model show higher flexibility in interpolating within the experimental points, thus the lower MAPE on GMM test points.
- The limitations of this study include the small scale of the training dataset, based only on 20 experimental points, partially addressed by the consideration of synthetic samples for the benchmark of the interpolation performance within the latent space.

In conclusion, and with respect to the FSP process considered in this research, in the presence of small datasets, the DNN formulation showed to be the most effective and robust modeling technique to model TS and WR. In this regard, it is expected that similar considerations may arise when carrying out similar benchmarks and analyses in different manufacturing processes, although the results trend and features to target variable relationships may change. To this end, as also highlighted previously, future work should explore hybrid strategies, such as physics-informed neural networks or expert-informed models, to further boost interpolation and especially extrapolation performance, enabling a better modeling of the FSP as well as other key manufacturing processes.

In summary, this work establishes a transparent and reproducible workflow for selecting and tuning predictive models in FSP/MMC applications under realistic, data-limited conditions and provides clear quantitative benchmarks to guide engineers and academicians in the choice of the best modeling strategy.

Appendix

Table 6 shows the results of the AIC (Akaike information criterion) and BIC (Bayesian information criterion) analyses from first to fourth polynomial regressions. Table 7 shows the percentage errors for the 20 cases of Table 3 for the case of employing second- and third-order regression functions for TS and WR.

Table 6 ANOVA results for TS and WR target variables

Metrics for TS (MPa)				
Polynomial order	R^2	AIC	BIC	# coefficients
1	0.6056	152.97	156.96	4
2	0.7474	156.06	166.02	10
3	0.9987	70.70	90.62	20
4	0.9987	100.70	135.55	35
Metrics for WR (mm^3/m)				
Polynomial order	R^2	AIC	BIC	# coefficients
1	0.3884	−18.28	−14.29	4
2	0.9267	−48.70	−38.75	10
3	0.9528	−37.50	−17.59	20
4	0.9528	−7.50	27.35	35

Table 7 ANOVA results for TS and WR target variables

	Percentage error for TS (%)		Percentage error for WR (%)	
Case#	Second order	Third order	Second order	Third order
1	4.78	0.00E+00	1.96	2.17E−14
2	10.55	5.80E−14	2.36	1.31E−14
3	17.87	8.70E−14	0.51	3.66E−14
4	3.63	1.07E−13	0.63	0.00E+00
5	2.13	1.18E−13	0.51	3.46E−14
6	5.43	9.19E−14	0.50	6.83E−14
7	5.73	3.66E−14	0.38	0.00E+00
8	0.29	0.00E+00	0.41	5.08E−14
9	24.38	4.35E−14	1.57	1.95E−14
10	13.03	1.67E−14	1.74	0.00E+00
11	14.19	5.74E−14	3.33	2.43E−14
12	26.84	1.28E−13	0.54	1.14E−14
13	25.32	2.58E−14	2.18	0.00E+00
14	1.04	3.06E−14	1.92	5.08E−14
15	3.29	5.14E−02	1.19	1.88E+00
16	2.08	1.19E+00	1.34	6.41E−01
17	1.46	1.83E+00	0.06	6.33E−01
18	4.47	1.27E+00	2.41	3.09E+00
19	4.76	1.57E+00	1.19	1.88E+00
20	3.29	5.14E−02	8.28	7.53E+00
Avg	8.72	2.99E−01	1.65	7.82E−01
Std. dev	8.39	5.96E−01	1.74	1.76E+00

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Vahid Modanloo and Majid Elyasi. Luca Quagliato created, ran the numerical analyses, and prepared the results. All authors were involved in analyzing the results and the discussion points for the paper. The first draft of the manuscript was written by Vahid Modanloo and Luca Quagliato, and all authors commented on previous versions of the manuscript. All authors read and approved of the final manuscript.

Funding Prof. Luca Quagliato was supported by the RP-Grant 2025 of Ewha Womans University.

Data availability The raw data related to this article and the algorithms developed by the authors are made available on request to the corresponding author. During the preparation of this work, no AI tool were employed for any text or content generation.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Samal P, Vundavilli PR, Meher A, Mahapatra MM (2020) Recent progress in aluminum metal matrix composites: a review on processing, mechanical and wear properties. J Manuf Process 59:131–152. https://doi.org/10.1016/j.jmapro.2020.09.010
2. Jiang J, Jiang F, Zhang M, Yi K (2023) Microstructure evolution and tensile property of deformed Al–Mg–Sc alloy: comparison of ECAP and FSP. J Mater Res Technol 22:2612–2626. https://doi.org/10.1016/j.jmrt.2022.12.109
3. Kar A, Sharma A, Kumar S (2024) A critical review on recent advancements in aluminium-based metal matrix composites. Crystals 14:. https://doi.org/10.3390/cryst14050412
4. Węglowski MS (2018) Friction stir processing – state of the art. Arch Civ Mech Eng 18:114–129. https://doi.org/10.1016/j.acme.2017.06.002
5. Sahraeinejad S, Izadi H, Haghshenas M, Gerlich AP (2015) Fabrication of metal matrix composites by friction stir processing with different particles and processing parameters. Mater Sci Eng A 626:505–513. https://doi.org/10.1016/j.msea.2014.12.077
6. Bharti S, Ghetiya ND, Patel KM (2021) A review on manufacturing the surface composites by friction stir processing. Mater Manuf Process 36:135–170. https://doi.org/10.1080/10426914.2020.1813897
7. Mehdi H, Mishra RS (2022) Consequence of reinforced SiC particles on microstructural and mechanical properties of AA6061 surface composites by multi-pass FSP. J Adhes Sci Technol 36:1279–1298. https://doi.org/10.1080/01694243.2021.1964846
8. Moustafa EB, Mikhaylovskaya AV, Taha MA, Mosleh AO (2022) Improvement of the microstructure and mechanical properties by hybridizing the surface of AA7075 by hexagonal boron nitride with carbide particles using the FSP process. J Mater Res Technol 17:1986–1999. https://doi.org/10.1016/j.jmrt.2022.01.150
9. Ahmadkhaniha D, Heydarzadeh Sohi M, Zarei-Hanzaki A et al (2015) Taguchi optimization of process parameters in friction stir processing of pure Mg. J Magnes Alloy 3:168–172. https://doi.org/10.1016/j.jma.2015.04.002
10. Das H, Mondal M, Hong ST et al (2018) Joining and fabrication of metal matrix composites by friction stir welding/processing. Int J Precis Eng Manuf - Green Technol 5:151–172. https://doi.org/10.1007/s40684-018-0016-7
11. Bajakke PA, Malik VR, Deshpande AS (2019) Particulate metal matrix composites and their fabrication via friction stir processing–a review. Mater Manuf Process 34:833–881. https://doi.org/10.1080/10426914.2019.1605181
12. Salah AN, Mabuwa S, Mehdi H et al (2022) Effect of multipass FSP on Si-rich TIG welded joint of dissimilar aluminum alloys AA8011-H14 and AA5083-H321: EBSD and microstructural evolutions. SILICON 14:9925–9941. https://doi.org/10.1007/s12633-022-01717-4
13. Thangarasu A, Murugan N, Dinaharan I, Vijay SJ (2015) Synthesis and characterization of titanium carbide particulate reinforced AA6082 aluminium alloy composites via friction stir processing. Arch Civ Mech Eng 15:324–334. https://doi.org/10.1016/j.acme.2014.05.010
14. Devaraju A, Kumar A, Kumaraswamy A, Kotiveerachari B (2013) Influence of reinforcements (SiC and Al2O3) and rotational speed on wear and mechanical properties of aluminum alloy 6061–T6 based surface hybrid composites produced via friction stir processing. Mater Des 51:331–341. https://doi.org/10.1016/j.matdes.2013.04.029
15. Sadeghi B, Shamanian M, Ashrafizadeh F, Cavaliere P (2017) FSW of bimodal reinforced Al-based composites produced via spark plasma sintering. Int J Mater Res 108:1045–1054. https://doi.org/10.3139/146.111573
16. Selvakumar S, Dinaharan I, Palanivel R, Ganesh Babu B (2017) Characterization of molybdenum particles reinforced Al6082 aluminum matrix composites with improved ductility produced using friction stir processing. Mater Charact 125:13–22. https://doi.org/10.1016/j.matchar.2017.01.016
17. Rahsepar M, Jarahimoghadam H (2016) The influence of multipass friction stir processing on the corrosion behavior and mechanical properties of zircon-reinforced Al metal matrix composites. Mater Sci Eng A 671:214–220. https://doi.org/10.1016/j.msea.2016.05.056
18. Sharma A, Fujii H, Paul J (2020) Influence of reinforcement incorporation approach on mechanical and tribological properties of AA6061- CNT nanocomposite fabricated via FSP. J Manuf Process 59:604–620. https://doi.org/10.1016/j.jmapro.2020.10.016
19. Wu B, Ibrahim MZ, Raja S et al (2022) The influence of reinforcement particles friction stir processing on microstructure, mechanical properties, tribological and corrosion behaviors: a review. J Mater Res Technol 20:1940–1975. https://doi.org/10.1016/j.jmrt.2022.07.172
20. Barmouz M, Givi MKB (2011) Fabrication of in situ Cu/SiC composites using multi-pass friction stir processing: evaluation of microstructural, porosity, mechanical and electrical behavior. Compos Part A Appl Sci Manuf 42:1445–1453. https://doi.org/10.1016/j.compositesa.2011.06.010
21. Yang Z, Wang J, Li X et al (2023) Improving the mechanical properties and electrical conductivity of cold-sprayed Cu-Ti3SiC2 composite by friction stir processing. Compos Part A Appl Sci Manuf 173:107698. https://doi.org/10.1016/j.compositesa.2023.107698
22. Merah N, Azeem MA, Abubaker HM, et al (2021) Friction stir processing influence on microstructure, mechanical, and corrosion behavior of steels: a review. Materials (Basel) 14:. https://doi.org/10.3390/ma14175023
23. Ding Z, Fan Q, Wang L (2019) A review on friction stir processing of titanium alloy: characterization, method, microstructure, properties. Metall Mater Trans B Process Metall Mater Process Sci 50:2134–2162. https://doi.org/10.1007/s11663-019-01634-9
24. Bagheri B, Abbasi M, Abdollahzadeh A, Kokabi AH (2020) A comparative study between friction stir processing and friction stir vibration processing to develop magnesium surface nanocomposites. Int J Miner Metall Mater 27:1133–1146. https://doi.org/10.1007/s12613-020-1993-4
25. Abbasi M, Givi M, Bagheri B (2019) Application of vibration to enhance efficiency of friction stir processing. Trans Nonferrous Met Soc China (English Ed 29:1393–1400. https://doi.org/10.1016/S1003-6326(19)65046-6
26. Barati M, Abbasi M, Abedini M (2019) The effects of friction stir processing and friction stir vibration processing on mechanical, wear and corrosion characteristics of Al6061/SiO2

surface composite. J Manuf Process 45:491–497. https://doi.org/10.1016/j.jmapro.2019.07.034
27. Abbasi M, Givi M, Ramazani A (2019) Friction stir vibration processing: a new method to improve the microstructure and mechanical properties of Al5052/SiC surface nanocomposite layer. Int J Adv Manuf Technol 100:1463–1473. https://doi.org/10.1007/s00170-018-2783-2
28. Abdollahzadeh A, Bagheri B, Abbasi M, et al (2021) Mechanical, wear and corrosion behaviors of AZ91/SiC composite layer fabricated by friction stir vibration processing. Surf Topogr Metrol Prop 9:. https://doi.org/10.1088/2051-672X/ac2176
29. Bagheri B, Abdollahzadeh A, Sharifi F, Abbasi M (2022) The role of vibration and pass number on microstructure and mechanical properties of AZ91/SiC composite layer during friction stir processing. Proc Inst Mech Eng Part C J Mech Eng Sci 236:2312–2326. https://doi.org/10.1177/09544062211024281
30. Men J, Paidar M, Eslami-Farsani R et al (2024) Investigation on microstructure, mechanical and tribological properties of friction stir processing of AZ31/AlFeCrMoNb surface composite. Mater Chem Phys 317:129149. https://doi.org/10.1016/j.matchemphys.2024.129149
31. Quagliato L, Seitz J, Perin M (2025) Machine learning modeling for material science and manufacturing : overview and perspectives for the future. 2025:1–1. https://doi.org/10.55092/aimat20250005
32. Xames MD, Torsha FK, Sarwar F (2023) A systematic literature review on recent trends of machine learning applications in additive manufacturing. J Intell Manuf 34:2529–2555. https://doi.org/10.1007/s10845-022-01957-6
33. Cao J, Bambach M, Merklein M et al (2024) Artificial intelligence in metal forming. CIRP Ann 73:561–587. https://doi.org/10.1016/j.cirp.2024.04.102
34. Verma S, Gupta M, Misra JP (2018) Performance evaluation of friction stir welding using machine learning approaches. MethodsX 5:1048–1058. https://doi.org/10.1016/j.mex.2018.09.002
35. Chadha U, Selvaraj SK, Gunreddy N, et al (2022) A survey of machine learning in friction stir welding, including unresolved issues and future research directions. Mater Des Process Commun 2022:. https://doi.org/10.1155/2022/2568347
36. Elsheikh AH (2023) Applications of machine learning in friction stir welding: Prediction of joint properties, real-time control and tool failure diagnosis. Eng Appl Artif Intell 121:105961. https://doi.org/10.1016/j.engappai.2023.105961
37. Pedrammehr S, Sajed M, Al-Abdullah KIAL, et al (2024) Experimental and machine learning study on friction stir surface alloying in Al1050-Cu alloy. J Manuf Mater Process 8:. https://doi.org/10.3390/jmmp8040163
38. Koch J, Choi WJ, King E et al (2024) Neural lumped parameter differential equations with application in friction-stir processing. J Intell Manuf 36:1111–1121. https://doi.org/10.1007/s10845-023-02271-5
39. Sadeghi B, Shamanian M, Cavaliere P et al (2018) Microstructural and mechanical behavior of bimodal reinforced Al-based composites produced by spark plasma sintering and FSP. Int J Adv Manuf Technol 94:3903–3916. https://doi.org/10.1007/s00170-017-1144-x
40. Lu CL, Chiu SY, Hsu CH, Yen SJ (2014) Enhanced differential evolution based on adaptive mutation and wrapper local search strategies for global optimization problems. J Appl Res Technol 12:1131–1143. https://doi.org/10.1016/S1665-6423(14)71672-4
41. Salgotra R, Gandomi AH (2024) A novel multi-hybrid differential evolution algorithm for optimization of frame structures. Sci Rep 14:1–28. https://doi.org/10.1038/s41598-024-54384-3
42. Tamilselvi S, Baskar S (2014) Covariance matrix adaptation evolutionary strategy for the solution of transformer design optimization problem. In: Panigrahi BK, Suganthan PN, Das S, Dash SS (eds) Swarm, Evolutionary, and Memetic Computing. SEMCCO 2013. Lecture Notes in Computer Science, vol 8297. Springer, Cham, pp 47–58. https://doi.org/10.1007/978-3-319-03753-0_5
43. Fu J, Sun X, Ma Z, et al (2024) Efficient optimization: unveiling the application of ensemble learning combined with the CMA-ES algorithm in hydraulic fracturing design. Processes 12:. https://doi.org/10.3390/pr12102299
44. Lee S, Park J, Kim N et al (2023) Extreme gradient boosting-inspired process optimization algorithm for manufacturing engineering applications. Mater Des 226:111625. https://doi.org/10.1016/j.matdes.2023.111625
45. Modanloo V, Jang S, Lee T, Quagliato L (2025) Gradient enhanced-expert informed neural network (GE-EINN) for forming depth prediction from a small-scale metal stamping dataset. J Manuf Process 140:224–240. https://doi.org/10.1016/j.jmapro.2025.02.052
46. Shimizu N, Kaneko H (2020) Direct inverse analysis based on Gaussian mixture regression for multiple objective variables in material design. Mater Des 196:109168. https://doi.org/10.1016/j.matdes.2020.109168
47. Quagliato L, Kim S, Hassan OR, Lee T (2025) Heel pad's hyperelastic properties and gait parameters reciprocal modelling by a Gaussian mixture model and extreme gradient boosting framework. Biomed Signal Process Control 107:107818. https://doi.org/10.1016/j.bspc.2025.107818
48. He X, Zhou Y (2018) Enhancing the performance of differential evolution with covariance matrix self-adaptation. Appl Soft Comput J 64:227–243. https://doi.org/10.1016/j.asoc.2017.11.050
49. Hansen N (2023) The CMA evolution strategy: a tutorial. Compiled March 13, 2023. Available at https://arxiv.org/abs/1604.00772
50. Bishop CM (2006) Chapter 9: mixture models and EM. In: Pattern recognition and machine learning. Springer, New York
51. Ling J, Hutchinson M, Antono E, Paradiso SP, Lookman T (2017) High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. Integr Mater Manuf Innov 6:207–217. https://doi.org/10.1007/s40192-017-0098-z
52. Laska A, Sadeghi B, Sadeghian B et al (2023) Temperature evolution, material flow, and resulting mechanical properties as a function of tool geometry during friction stir welding of AA6082. J Mater Eng Perform 32:10655–10668. https://doi.org/10.1007/s11665-023-08671-1
53. Dolatkhah A, Golbabaei P, Besharati Givi MK, Molaiekiya F (2012) Investigating effects of process parameters on microstructural and mechanical properties of Al5052/SiC metal matrix composite fabricated via friction stir processing. Mater Des 37:458–464. https://doi.org/10.1016/j.matdes.2011.09.035
54. Rajwar K, Deep K, Das S (2023) An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges. Artif Intell Rev 56:13187–13257. https://doi.org/10.1007/s10462-023-10470-y
55. Tomar V, Bansal M, Singh P (2023) Metaheuristic algorithms for optimization: a brief review. Eng Proc 59:1–16. https://doi.org/10.3390/engproc2023059238
56. Bajaj NS, Patange AD, Jegadeeshwaran R et al (2023) Application of metaheuristic optimization based support vector machine for milling cutter health monitoring. Intell Syst with Appl 18:200196. https://doi.org/10.1016/j.iswa.2023.200196
57. Montgomery DC (2017) Design and analysis of experiments, 9th edn. John Wiley & Sons, Hoboken
58. Mirandola I, Berti GA, Caracciolo R, et al (2021) Multivariable regression and gradient boosting algorithms for energy prediction in the radial-axial ring rolling (rarr) process. ESAFORM 2021 - 24th Int Conf Mater Form 04:1–10. https://doi.org/10.25518/esaform21.3775