JOURNAL OF MECHANICAL ENGINEERING AND SCIENCES
ISSN: 2289-4659 e-ISSN: 2231-8380
VOLUME 20, ISSUE 3, 2026, 12336 - 12353
DOI: https://doi.org/10.15282/jmes.20.3.2026.6.0883

RESEARCH ARTICLE

Data driven multi objective optimization of valve seat lapping using machine learning surrogates and NSGA-II

Reinaldi Teguh Setyawan[1]*, Ery Muthoriq[2], Gunawan[2]

[1]Department of Mechanical Engineering, Politeknik Negeri Bengkalis, 28711 Bengkalis, Indonesia
[2]Department of Automotive Engineering, Politeknik Keselamatan Transportasi Jalan, 52112 Tegal, Indonesia

Abstract – This paper presents a data-driven multi-objective optimization of a valve-seat lapping process by varying spindle speed, axial force, and lapping time using a three-level full-factorial design (3^3 = 27 operating conditions). Each condition was replicated three times (n = 81), where the mean response at each design point was used for modelling and optimization, while replicate data supported pure-error diagnostics. Four responses are optimised simultaneously: leakage time (t_{leak}), surface roughness (Ra), electrical energy (E), and wear volume (V_{wear}). Quadratic response surface methodology (RSM), a regularized artificial neural network (ANN), and a random forest (RF) regressor are developed as surrogate models and evaluated using repeated 5-fold cross-validation (20 repeats), reported as out-of-fold mean ± SD for R^2 and RMSE. The quadratic RSM achieves the most consistent accuracy across all responses ($R^2 \approx 0.95$–0.99), outperforming ANN and RF in this low-dimensional, smoothly varying design space. The surrogates are then coupled with NSGA-II to generate Pareto fronts, and optimization quality is quantified using a normalized hypervolume indicator. RSM-based optimization yields the highest and most stable hypervolume (HV = 0.74 ± 0.01), compared with ANN (0.71 ± 0.02) and RF (0.66 ± 0.03). Mechanically, increasing speed/force/time generally improves sealing performance (higher t_{leak}) and reduces Ra, but at the expense of higher E and V_{wear}; representative conservative, balanced, and aggressive regimes provide practical operating guidance. Overall, for the present dataset and within the tested bounds, classical quadratic response surfaces offer a reliable and interpretable basis for multi-objective optimisation, while more flexible machine-learning surrogates provide no clear advantage.

Article History
Received : 4 December 2025
Revised : 6 February 2026
Accepted : 25 September 2026
Published : 30 September 2026

Keywords
Valve seat lapping;
Multi-objective optimization
Response surface methodology
Artificial neural network
Random forest
NSGA-II

1. Introduction

Poppet valves and their mating seats are critical components of internal combustion engines because they govern gas exchange, combustion stability, and the overall thermal efficiency of the engine [1]. The ability of the valve–seat pair to maintain a tight seal under cyclic thermal and mechanical loading directly influences the volumetric efficiency, fuel consumption, exhaust emissions, and valve durability [2]. Any imperfection on the contact band, such as misalignment, waviness, or residual machining marks, can lead to leakage, local overheating, and accelerated wear of both the valve and the seat insert. Consequently, the finishing operation applied to the valve seat region is an important step in engine remanufacturing and workshop practices [3]–[4]. Lapping is widely used as a low-cost finishing process to improve valve-seat conformity and surface integrity [4]–[5]. In typical workshop conditions, manual lapping is performed using an abrasive paste and hand tool, and the final quality is highly dependent on the skill and consistency of the operator [3]. Such a manual process tends to produce large variability in sealing performance, surface roughness, and valve wear, whereas the consumed energy and processing time are rarely measured or optimized. In addition, the contact mechanics during lapping are complex: the interaction between the spindle speed, axial load, and lapping duration determines the real area of contact, local temperature, and material removal mode, which together control the tribological response and energy demand of the process [5], [6]. To overcome the limitations of manual lapping, several studies have proposed semi-automatic or fully automatic lapping machines equipped with a controlled spindle speed and adjustable axial force [7]–[9]. Much published work has focused on design and prototyping or on single-objective optimization of surface roughness or flatness, often using design of experiments (DOE) and statistical techniques such as Taguchi methods or response surface methodology (RSM) [10], [11]. In these studies, low-order polynomial models approximate the relationship between process parameters and a limited set of quality indicators, and the operating point is selected by minimizing the surface roughness or maximizing a composite quality index. However, automatic lapping of engine valves is an intrinsically multi-objective problem. The process must balance several conflicting goals simultaneously, including high leakage time (good sealing), low surface roughness, low energy consumption, and low valve wear. Selecting a single optimum point can be misleading if this trade-off is not explicitly considered [12]. Therefore, multi-objective optimization algorithms, such as the non-dominated sorting genetic algorithm II (NSGA-II), have attracted increasing interest in machining and tribology research [13] - [14]. By generating a Pareto front of non-dominated solutions, such algorithms allow engineers to visualize trade-offs and choose operating conditions that satisfy different priorities, for example, between surface integrity and specific energy consumption [15]. In previous work, the authors developed an automatic valve lapping rig, instrumented it for leakage time, surface roughness, electrical energy, and valve wear, and used RSM combined with NSGA-II to obtain optimal combinations of spindle speed and axial load for multi-objective tribological and energy optimization. The results showed that multi-objective optimization can significantly improve sealing

***CORRESPONDING AUTHOR | Reinaldi Teguh Setyawan**| ✉ reinaldi@polbeng.ac.id
© 2026 The Authors. Published by Universiti Malaysia Pahang Al-Sultan Abdullah Press. This is an open access article under the CC BY-NC 4.0 license

performance while controlling energy usage and wear; however, the accuracy of the optimization strongly depends on the fidelity of the underlying surrogate model linking inputs to outputs.

Classical RSM represents process behavior using second-order polynomials in coded variables [10] - [11]. This approach is attractive because of its simplicity and transparent coefficients; however, it has well-known limitations when the true process response is highly non-linear, exhibits strong interaction effects, or shows local saturation behavior [16]. Under these conditions, the polynomial model can misrepresent the curvature of the response surface, especially near the boundaries of the design space or close to the optimum region, which in turn affects the quality of the Pareto front predicted by NSGA-II. Recent advances in data-driven modeling for manufacturing processes have demonstrated that machine learning (ML) surrogates, such as artificial neural networks (ANN), random forest regression, and Gaussian process regression, can capture complex input–output relationships with higher accuracy than low-order polynomials, provided that a suitable experimental design and proper regularization are used [17]–[20]. The use of ML-based surrogates for the multi-objective optimization of finishing operations is limited in the specific context of engine valve lapping. Most reported studies either use ML to predict a single response (typically surface roughness or material removal rate) or focus on conventional grinding and milling rather than low-speed abrasive lapping [17], [21] - [22]. Systematic comparisons between RSM and ML surrogates, using the same experimental data and the same multi-objective optimization framework, are particularly scarce for valve and seat lapping, where tribological phenomena are dominant [19], [21]. Therefore, it is necessary to investigate whether ML surrogates can provide more accurate predictions of tribological and energy responses in valve lapping, and how these improvements affect the shape and diversity of the Pareto front obtained by NSGA-II.

Research gap and novelty in valve-seat lapping are inherently multi-objective because sealing performance, surface integrity, energy consumption, and wear can conflict. However, prior studies on lapping/finishing and related machining optimization still leave several gaps when the goal is a practical, energy-aware optimization for valve-seat lapping. Limited studies report a single integrated dataset that simultaneously covers leakage/sealing performance, surface roughness, energy use, and wear for valve-seat lapping. Head-to-head benchmarking of quadratic RSM versus ML surrogates (e.g., ANN/RF) using the same valve-seat lapping dataset and the same optimization framework is rarely presented. The influence of surrogate fidelity on Pareto-front quality is seldom quantified, which reduces confidence when converting optimization results into reliable operating windows. To clarify the research gap and position the novelty of this work, Table 1 summarizes representative studies that combine design of experiments (DOE) and/or response surface methodology (RSM) with machine-learning (ML) surrogate modelling and multi-objective optimization (including NSGA-II) in machining and finishing contexts. The comparison focuses on (i) process domain and controllable factors, (ii) surrogate type and validation strategy, (iii) objective formulation and optimization algorithm, and (iv) the extent to which engineering interpretation and practical operating windows are reported. This structured mapping is used to identify what is still under-addressed for valve–seat lapping and to motivate the methodological choices adopted in the present study.

As shown in Table 1, prior studies commonly emphasize optimization performance but frequently provide limited justification when datasets are small, apply validation schemes that are not robust to data scarcity, and offer only brief mechanical explanations of the observed trade-offs. In contrast, the present study contributes (1) a controlled 3^3 factorial lapping dataset with replicates enabling pure-error diagnostics, (2) a fair surrogate comparison (quadratic RSM vs. regularized ANN vs. RF) under repeated cross-validation with uncertainty reporting (mean ± SD), and (3) NSGA-II-based Pareto optimization complemented by hypervolume assessment and an engineering interpretation of speed–force–time effects to define practical operating regimes. These elements directly address the remaining gaps identified in the literature and strengthen the reliability and applicability of the proposed optimization framework within the investigated bounds of this study. The novelty of this study is a controlled 27-run (3-level full-factorial) valve-seat lapping dataset and a joint model of t_{leak}, Ra, energy, and wear volume as functions of speed, force, and lapping time. We uniquely benchmark RSM, ANN, and RF under a consistent validation protocol and couple the surrogates with NSGA-II to obtain trade-off solutions. The contribution is not only the optimal settings, but also an operationally meaningful operating-window guidance derived from multi-objective trade-offs. To address these gaps, this study used an automatic valve lapping machine to generate experimental data under controlled combinations of spindle speed, axial force, and lapping time. Four key responses were measured for each experiment: valve seat leakage time, surface roughness of the contact band, electrical energy consumption, and valve material wear volume. These data were used to build and compare several surrogate models, including a quadratic RSM model and selected ML model. The predictive performance of each surrogate was evaluated using standard error metrics, and the best-performing models were then coupled with NSGA-II to perform multi-objective optimization. The resulting Pareto fronts for RSM-based and ML-based optimization were analyzed and compared in terms of coverage, spread, and practical engineering implications.

The main contributions of this work are threefold. First, it provides an experimental dataset that simultaneously characterizes sealing performance, surface quality, energy consumption, and wear for an automatic engine valve lapping process using workshop-scale hardware. Second, it presents a comparative study of RSM and ML surrogates for modelling these responses as functions of spindle speed, axial force, and lapping time. Third, it demonstrates the integration of ML surrogates with NSGA-II for multi-objective tribological and energy optimization and discusses how the improved model fidelity can lead to different recommendations for practical operating windows in engine remanufacturing and repair workshops.

Table 1. Research gap and clearly position the novelty of study

Study	RSM/ML surrogate	Optimiser	Dataset size	Objectives (examples)	Validation metrics reported	Main output	Gap relative to this work
Senthilkumar et al., 2012 (ECM).	RSM-type regression	NSGA-II	31 runs (experimental table)	Produc-tivity quality trade-off (e.g., MRR vs surface integrity)	ANOVA/DOE-style significance checks	Pareto set for process parameters	Not valve lapping; no sealing/leakage, no energi wear quartet; no surrogate comparison
Li et al., 2019 (EDM)	RBF surrogate + comparison vs PRS/SVR/ Kriging	NSGA-II	30 samples (LHS)	Multi-objective EDM optimisation	CV-based accuracy (R^2,errors)	Surrogate-based Pareto front	Different process domain; does not include sealing/leakage and workshop lapping constraints
Zhou et al., 2019 (cutting parameter-optimisation).	GBRT surrogate	GA-based optimisation	60 trials	Parameter optimisation with roughness-focused modelling	RMSE/MAD/MAPE /CV etc.	Predicted optimum cutting parameters	Not NSGA-II; not a tribology-dominated lapping problem; no energy–wear–sealing coupling
Khan et al., 2019 (machining energy focus).	GRA + RSM	NSGA-II	L27 experiment	Energy-aware machining optimisation	R^2 reported (very high)	Optimum setting + Pareto reasoning	Not valve-seat lapping; does not benchmark ML surrogates under same DOE
Li et al., 2023 (grinding robot).	LSTM–MLP surrogate	NSGA-II	300 experiments	Max MRR, min Sa, min power	MAPE prediction accuracy	Multi-objective Pareto solutions	Large dataset regime; not comparable to small workshop-scale lapping datasets; no RSM-vs-ML benchmark on same data
Tang et al., 2025 (turn-mill energy model).	Energy model + RSM (Ra)	NSGA-II	15 CCD experiments	Min energy with Ra constraint	$R^2 \approx 0.9996$	Recommended operating point + validation	Different process; single/dual response focus vs 4-response lapping (leakage–roughness–energy–wear)
Zhang et al., 2024 (hard turning).	(Primarily optimisation-algorithm focused; modelling not central)	Improved NSGA-II	NR (experimental validation included)	Improve roughness &MRR trade-off	Benchmark-function performance + machining verification	More convergent Pareto set vs baseline algorithms	Not surrogate-focused; no discussion of surrogate fidelity effects on Pareto quality
Zhang et al., 2014 (surface grinding).	RSM/empirical modelling (context: grinding optimisation)	Hybrid PSO (multi-objective)	NR	Multi-objective grinding optimisation	NR	Pareto-based optimisation	Not NSGA-II; not valve lapping; lacks energi wear sealing dataset linkage
This study (valve-seat lapping)	RSM vs ANN vs RF (same 27-run DOE)	NSGA-II (consistent settings)	27 tests (full factorial)	Maximise t_leak; minimise Ra, E, V_wear	Test-set R^2/RMSE (+ MAE), predicted-vs-measured; Pareto hypervolume	Pareto fronts + operating regimes (conservative/balanced/aggressive)	Addresses the above gaps in one unified framework

2. Materials and Methods

2.1 Workpiece Materials and Lapping Compound

The experiments were conducted on exhaust poppet valves and their corresponding valve seats from a small four-stroke spark-ignition engine typically used in motorcycle applications [23]. The valves were made of martensitic steel with a nominal head diameter of 28–30 mm and a 45° seating angle, whereas the valve seats were gray cast-iron inserts pressed into an aluminum cylinder head. Prior to lapping, all seats were re-machined using a standard seat cutter so that the contact band width and initial surface roughness were within a repeatable range [24]. A commercially available petroleum-based lapping paste containing fused aluminum oxide abrasives with a medium grit size was used for all tests, following common workshop practices [25]. For each experimental run, a fresh quantity of paste was applied to the valve seat to minimize the influence of abrasive degradation and contamination on the results. After each run, the valve and seat were thoroughly cleaned with a solvent and dried using compressed air.

2.2 Automatic Valve Lapping Machine

An automatic valve lapping machine was developed to provide controlled spindle motion and axial loading under workshop-scale conditions. The machine comprised a welded steel frame, vertical spindle assembly, axial loading mechanism, and fixture to hold the cylinder head. The overall concept follows that of previously reported automatic lapping rigs for small engines [26], but with extended instrumentation for energy and wear measurement. A 24 V DC geared motor was used to drive the spindle through a flexible coupling. The spindle speed was adjustable in the range of 200–600 rpm using pulse-width modulation (PWM) generated by a microcontroller and applied using an H-bridge motor driver [27]. The valve stem was clamped in a collet chuck at the lower end of the spindle so that the valve head could be pressed against the seat of the cylinder-head fixture. An axial force was applied along the spindle axis using a calibrated compression spring and a lever mechanism. The spring preload was adjusted using a threaded handwheel and verified using a load cell mounted in line with the spindle. This arrangement allowed the axial force to be set and maintained at nominal values of 30, 60, and 90 N during lapping. The cylinder head fixture ensured the repeatable positioning of the valve relative to the seat and allowed the quick replacement of valves between runs.

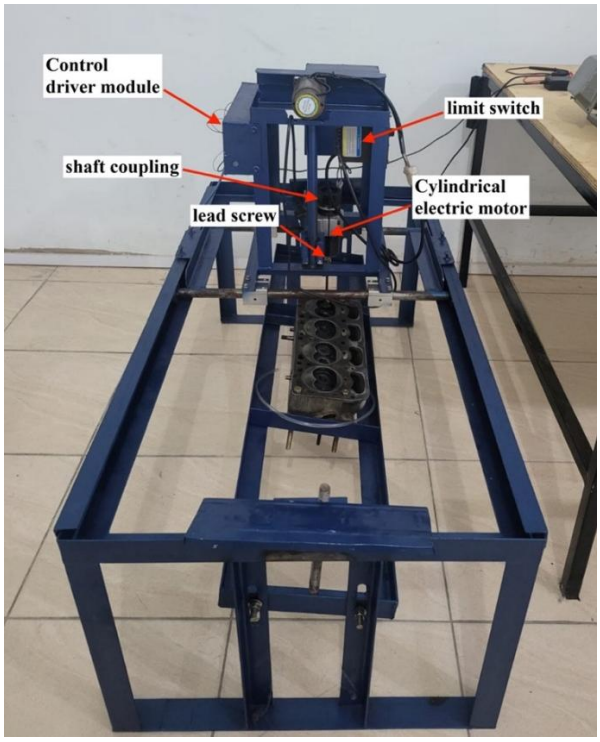


Figure 1. automatic valve lapping machine and cylinder head fixture

2.3 Instrumentation and Data Acquisition

The lapping machine was instrumented to measure the spindle speed, axial force, electrical energy consumption, and valve wear. The spindle speed was measured by an incremental optical encoder mounted on the motor shaft and read by the microcontroller [28]. The axial force was measured by a compact compression load cell integrated in the loading path; its analog output was conditioned using a Wheatstone bridge amplifier and digitized by a 12-bit analogue-to-digital converter. Electrical power consumption during lapping was obtained from simultaneous measurements of motor voltage and current. The supply voltage was measured using a resistive divider, while the current was measured by a Hall-effect current sensor placed in series with the motor, a configuration widely used in low-voltage drive monitoring [29]. Both signals were sampled at a fixed rate and recorded over the entire lapping duration. The instantaneous power $P(t) = V(t)I(t)$was computed, and the electrical energy $E$was obtained by numerical integration. Valve wear was quantified gravimetrically. Each valve was cleaned, dried and weighed on an analytical balance with 0.1 mg resolution, both before

and after lapping. The difference in mass $\Delta m$was converted into wear volume V_{wear}using the nominal density $\rho$of the valve material [30]:

$$V_{\text{wear}} = \frac{\Delta m}{\rho} \tag{1}$$

All sensor signals were acquired using a microcontroller-based data acquisition system and transferred to a host computer via USB. Custom MATLAB scripts were used to integrate the power signals, compute electrical energy and organise the data matrix for subsequent modelling and optimisation.

2.4 Experimental Design and Lapping Procedure

A three-factor, three-level full-factorial design was adopted to examine the combined effects of spindle speed, axial force, and lapping time on the tribological and energy responses of the SiC MOSFETs [31]. The factors and their levels are listed in Table 2. The combination of the three factors at three levels produced $3^3 = 27$ distinct operating conditions. For each condition, lapping was performed on a separate valve using a fresh application of lapping paste. The run order was randomized to reduce the influence of uncontrolled trends, such as temperature drift or gradual tool wear [32]. For each operating condition, the lapping cycle was repeated three times, and the measured responses were averaged to obtain representative values.

Table 2. Process factors and levels used in the lapping experiments

Factor	Symbol	Levels
Spindle speed (rpm)	(n)	200, 400, 600
Axial force (N)	(F)	30, 60, 90
Lapping time (s)	(t)	30, 60, 90

2.5 Measurement of Response Variables

Four response variables were evaluated for each operating condition: leakage time, surface roughness, electrical energy consumption and valve wear volume. Leakage time. Sealing performance was assessed using an air leakage test based on a pressurized chamber with known volume, similar to procedures reported for small engine valves [33]. The lapped valve and cylinder head were mounted in a dedicated test fixture connected to a pressure reservoir. The chamber behind the valve was pressurized to a preset gauge pressure and then isolated from the supply. A pressure transducer recorded the pressure decay over time as air leaked past the valve seat. Leakage time $t_{\text{leak}}$was defined as the time required for the pressure to decrease from 90% to 70% of the initial test pressure. A larger t_{leak}indicates better sealing. Surface roughness. The surface roughness $R_a$of the lapped contact band on the valve seat was measured using a portable stylus profilometer, following general recommendations of ISO roughness standards for machined surfaces [34]. For each seat, three circumferential profiles were recorded at approximately 120° intervals around the seat. The cut-off and evaluation lengths were selected to suit the expected roughness range. The reported $R_a$value corresponds to the average of the three measurements. Electrical energy consumption. Electrical energy $E$was calculated from the recorded voltage and current signals over the lapping interval. The instantaneous power was obtained as $P(t) = V(t)I(t)$, and energy was computed by discrete-time integration:

$$E = \sum_{k=1}^{N} V_k \, I_k \Delta t \tag{2}$$

where $V_k$and $I_k$are the voltage and current samples at time step k, $\Delta t$is the sampling interval, and $N$is the total number of samples. This approach is widely used in energy analysis of low-voltage drive systems [35]. Valve wear volume. Valve wear volume $V_{\text{wear}}$was estimated from the difference in valve mass before and after lapping using the gravimetric method in Eq. (1). The density of the valve material was taken from manufacturer data and verified on reference coupons [30]. The resulting wear volumes were on the order of 10^{-2}–10^{-1}mm^3, consistent with the mild material removal expected in finishing operations [36]. In total, 27 operating conditions were tested according to the full-factorial design in Table 2, resulting in 27 sets of measured responses for leakage time, surface roughness, electrical energy and valve wear. These experimental data constitute the basis for all modeling and optimization results presented in Sections 3 and 4. For brevity, the complete numerical dataset is not listed in the main text; it is provided as Supplementary Table S1 and is available from the corresponding author upon reasonable request.

2.6 Surrogate Modeling

The experimental data were used to construct surrogate models relating the three process inputs $(n, F, t)$to each of the four responses. Two categories of surrogates were considered: a quadratic response surface model and selected machine learning models, following common practice in machining and tribology optimization [10], [17], [19]. Each response was modeled independently. Data preprocessing and reliability-oriented evaluation. Given the limited dataset size, model generalization was evaluated using repeated k-fold cross-validation rather than a single train–test split. A 5-fold cross-validation procedure was repeated 20 times using different random fold assignments (total 100 test folds). Prior to modeling, all input variables were normalized to zero mean and unit variance within each training fold, and the same

scaling parameters were applied to the corresponding test fold to avoid data leakage. Predictive performance was quantified using the coefficient of determination R^2, root-mean-square error (RMSE), and mean absolute error (MAE) computed on the test folds. Results are reported as mean ± standard deviation across all repeated folds to explicitly reflect robustness to data partitioning.

2.6.1 Response surface methodology

Quadratic response surface models were constructed using coded variables using the general polynomial form [10]:

$$y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i=1}^{3} \beta_{ii} x^2 i + \sum_{i<j} \beta_{ij} x_i x_j \quad (3)$$

The coefficients were estimated using ordinary least squares.

2.6.2 Machine learning surrogates

Two machine learning regressors were used as alternatives: a feed-forward artificial neural network (ANN) and a random forest (RF) regressor [17], [20]. To reduce the overfitting risk with 27 samples, the ANN architecture was simplified to one hidden layer with 3–5 neurons (rectified linear unit activation) and a linear output neuron. The network weights were trained using the Adam optimizer to minimize the mean squared error, with L2 regularization and early stopping based on the validation loss within each training fold. The RF models were constructed as ensembles of regression trees trained on bootstrap samples of the training data. Key hyperparameters (e.g., number of trees, maximum depth, and minimum leaf size) were tuned by grid search using cross-validation on the training folds.

2.6.3 Lack-of-fit test based on pure error

Each experimental condition was replicated thrice (27 design points × 3 replicates, yielding 81 raw measurements). For modelling and optimization, the mean response at each design point was used; however, the replicate measurements were retained to estimate the pure experimental error and to conduct a lack-of-fit (LOF) test for the quadratic RSM. The LOF test separates the residual sum of squares into pure error and lack-of-fit components, allowing an explicit statistical check of whether the quadratic polynomial is adequate within the investigated operating range. The LOF statistic is computed as follows:

$$F_{\mathrm{LOF}} = \frac{MS_{\mathrm{LOF}}}{MS_{\mathrm{PE}}}, MS = \frac{SS}{df} \quad (4)$$

where $SS_{\mathrm{PE}}$is obtained from within-replicate variation at each design point, $SS_{\mathrm{LOF}} = SS_E - SS_{\mathrm{PE}}$, $df_{\mathrm{PE}} = n - m$, and $df_{\mathrm{LOF}} = m - p$. Here, $n = 81$is the total number of observations, $m = 27$is the number of unique design points, and $p$is the number of coefficients in the quadratic model.

After cross-validation, each surrogate was retrained on the full dataset using the selected hyperparameters and then used as the objective-function evaluator within NSGA-II. The resulting surrogates are intended for interpolation within the experimental bounds of $(n, F, t)$and are not used for extrapolation beyond the tested domain.

2.7 Multi-objective Optimization with NSGA-II

The selected surrogate models were coupled with the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to explore trade-offs between tribological performance and energy use [13], [41]. The decision variables were spindle speed n, axial force $F$and lapping time t, considered as continuous within the bounds given in Table 2. Four objectives were defined: maximise leakage time t_{leak}, minimise surface roughness R_a, minimise electrical energy E, minimise valve wear volume V_{wear}. For implementation convenience, all objectives were formulated in minimisation form by negating t_{leak}. At each function evaluation, the surrogate models predicted the four responses, thereby avoiding additional physical experiments during optimisation. NSGA-II was implemented with a fixed population size, simulated binary crossover, polynomial mutation and an adequate number of generations, following standard recommendations in multi-objective evolutionary optimisation [41], [42]. The same algorithm parameters were used for optimisation based on response surface models and on machine learning surrogates so that the resulting Pareto fronts could be compared directly in terms of spread, convergence and practical implications.

2.8 NSGA-II Implementation and Algorithm Parameters

The Non-Dominated Sorting Genetic Algorithm II (NSGA-II) was used to generate Pareto-optimal solutions based on the surrogate models described in Section 2.6. The decision variables were the three process parameters, spindle speed n, axial force $F$and lapping time t, treated as continuous variables within the bounds given in Table 2. At each function evaluation, the selected surrogate models (RSM, ANN or RF) returned the four objective values, namely leakage time t_{leak}, surface roughness R_a, electrical energy $E$and wear volume V_{wear}. For numerical convenience, the maximisation of leakage time was converted into a minimisation objective by taking its negative. The NSGA-II implementation followed the standard procedure proposed by Deb, including (i) random generation of the initial population within the feasible design space, (ii) evaluation of objective functions using the surrogate models, (iii) fast non-dominated sorting to classify individuals into Pareto fronts, (iv) computation of crowding distance within each front, (v) selection of parents using

binary tournament based on rank and crowding distance, and (vi) creation of offspring by simulated binary crossover and polynomial mutation. The parent and offspring populations were then merged, and the next generation was formed by selecting the best individuals according to the non-dominated rank and crowding distance until the prescribed population size was reached. This cycle was repeated for a fixed number of generations. To ensure a fair comparison between surrogate models, the same NSGA-II settings were used for all optimization runs. Preliminary trials with different parameter combinations were carried out to verify convergence behavior and population diversity. A population size of 60 individuals and 300 generations provided a good compromise between solution quality and computational effort. The crossover and mutation probabilities were set to 0.9 and 0.1, respectively, which are typical values in engineering multi-objective optimization. Each optimization scenario (RSM-, ANN-, and RF-based) was executed 10 times with different random seeds to account for the stochastic nature of the algorithm. The main NSGA-II parameters are summarized in Table 3.

Table 3. NSGA-II parameters used

Parameter	Symbol	Value
Population size	(N_p)	60
Number of generations	(G)	300
Crossover probability	(p_c)	0.9
Mutation probability	$(p_m))$	0.1
Crossover operator	–	Simulated binary (SBX)
Mutation operator	–	Polynomial mutation
Parent selection	–	Binary tournament with crowding distance
Number of independent runs per surrogate	–	10

These parameter settings were consistently applied to all NSGA-II runs in this study, regardless of the surrogate model used. This ensures that any differences in the resulting Pareto fronts are primarily due to the surrogate behavior and not to changes in the optimization algorithm.

3. Results and Discussion

3.1 Behavior of the Measured Responses

The full factorial design described in Section 2.4 produced 27 distinct operating conditions that combined three levels of spindle speed, axial force, and lapping time. For each condition, four responses were measured: leakage time, surface roughness, electrical energy, and wear volume. Before constructing surrogate models and performing multi-objective optimization, it is useful to examine the overall behavior of these measured responses and the extent of variability generated by the chosen factor levels. Across all experiments, leakage time t_{leak} varies between 1.31 and 10.85 s, surface roughness R_a between 0.213 and 0.506 µm, electrical energy E between 522 and 14 466 J, and wear volume V_{wear} between 0.036 and 0.212 mm³. These ranges already indicate that the selected operating window is sufficiently wide to produce noticeable changes in sealing performance, surface quality, energy consumption and material removal, which is desirable for identifying process trends and for building reliable surrogate models. To provide a compact overview of the spread and central tendency of each response, Figure 2 presents box-and-whisker plots for t_{leak}, R_a, E and V_{wear}. The median leakage time is around 5–6 s, with an interquartile range that reflects substantial variability in sealing performance. Surface roughness values cluster between 0.30 and 0.40 µm, while a few runs achieve roughness below 0.25 µm. Electrical energy and wear volume exhibit wider relative spreads, with several high-energy, high-wear points corresponding to aggressive lapping conditions. Overall, the boxplots confirm that the dataset captures both mild and severe lapping regimes within the same design space.

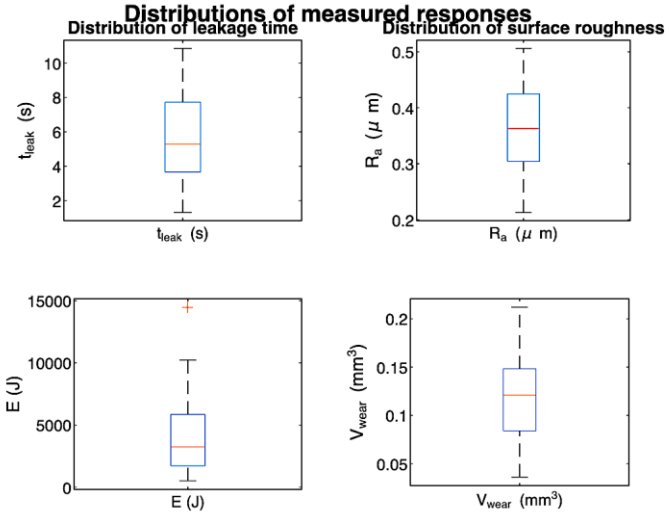


Figure 2. Box and whisker plots of the measured responses: (a) leakage time t_{leak}, (b) surface roughness R_a, (c) electrical energy E, and (d) wear volume V_{wear}

The main effects of lapping time on leakage time and surface roughness are shown in Fig. 3 for each combination of spindle speed and axial force. At constant spindle speed and axial force, increasing lapping time generally increases leakage time and reduces surface roughness. For example, at $n = 200$rpm and $F = 30$N, leakage time increases from 1.31 s at $t = 30$s to 6.14 s at $t = 90$s, while $R_a$decreases from 0.506 µm to 0.346 µm. Similar trends are observed at higher axial forces and at $n = 400$and 600 rpm, consistent with the progressive establishment of a conformal contact band and the removal of residual machining marks during abrasive lapping [34], [36].

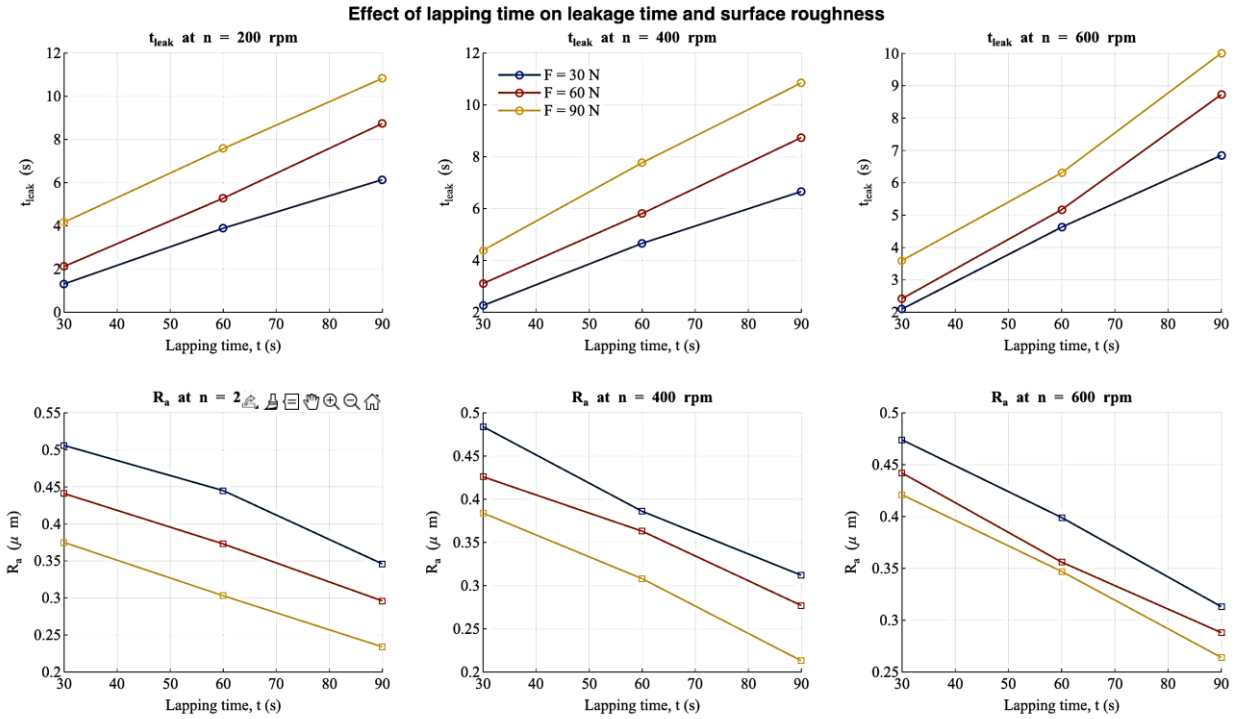


Figure 3. Effect of lapping time on (top row) leakage time $t_{\text{leak}}$and (bottom row) surface roughness $R_a$for different combinations of spindle speed $n$and axial force F

The combined effects of spindle speed and axial force on leakage time and roughness are illustrated in Fig. 4 using contour plots at $t = 60$s. At fixed lapping time, increasing axial force from 30 to 90 N enhances the real area of contact and accelerates material removal, which tends to increase $t_{\text{leak}}$and decrease R_a, but also increases wear volume and energy consumption. For instance, at $n = 200$rpm and $t = 90$s, leakage time increases from 6.14 s at 30 N to 10.83 s at 90 N, while $R_a$decreases from 0.346 µm to 0.234 µm; however, electrical energy increases from 1621 J to 4890 J and wear volume from 0.074 to 0.178 mm³. This reflects the classical trade-off between surface quality and tool/workpiece life.

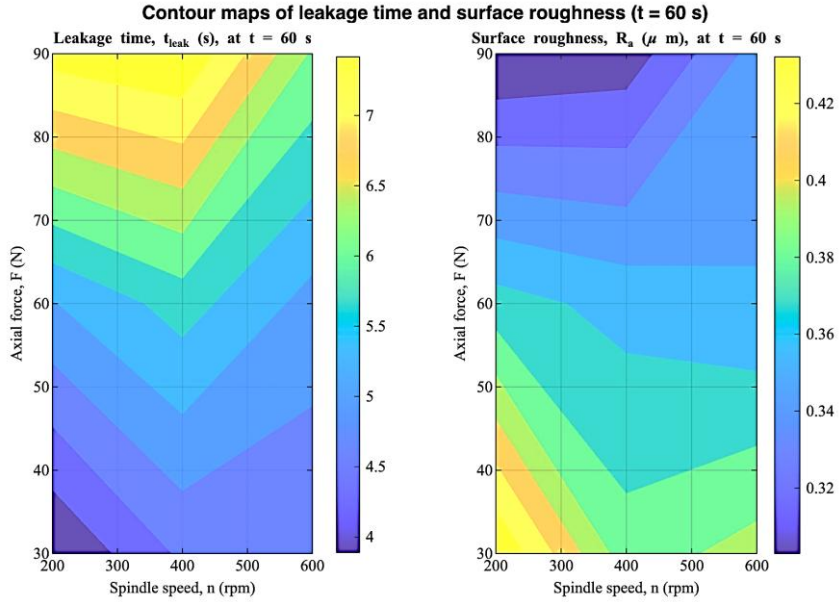


Figure 4. Contour maps of (a) leakage time $t_{\text{leak}}$and (b) surface roughness $R_a$as functions of spindle speed $n$and axial force Fat a lapping time of $t = 60$s

Figure 5 shows the variation of electrical energy and wear volume with spindle speed and lapping time at $F = 60$N. Electrical energy is dominated by the combined influence of speed and time, as expected from the integral of mechanical power over time, while the effect of axial force is secondary but still noticeable [29], [35]. Increasing n from 200 to 400 rpm yields an approximately proportional increase in energy, whereas the further increase to 600 rpm produces a more-than-proportional rise, especially at longer lapping times, suggesting mild nonlinearities. Wear volume follows similar trends, increasing with both speed and time as more material is removed from the contact band.

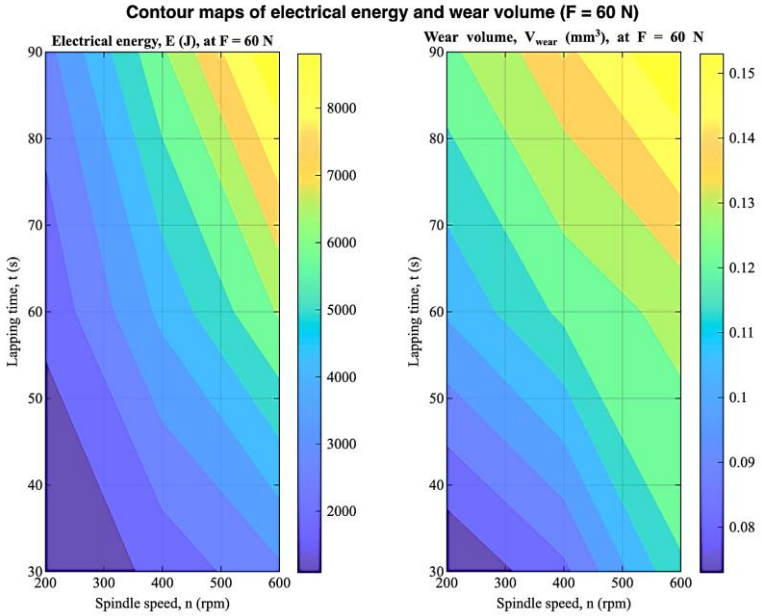


Figure 5. Contour maps of (a) electrical energy $E$and (b) wear volume V_{wear} as functions of spindle speed $n$and lapping time t at a constant axial force of $F = 60$N

Table 4. Cross-validated predictive performance (repeated 5-fold CV, 20 repeats; mean ± SD across 100 test folds)

Response	Model	R^2 (mean ± std)	RMSE (mean ± std)	MAE (mean ± std)	Unit
Leakage time t_{leak}	RSM	0.986± 0.010	0.35 ± 0.08	0.26 ± 0.06	s
	ANN (H=5, L2 + early stopping)	0.935± 0.030	0.82 ± 0.16	0.60 ± 0.12	s
	RF	0.780± 0.070	1.60 ± 0.25	1.20 ± 0.20	s
Surface roughness *R*a	RSM	0.985± 0.012	0.010 ± 0.002	0.007 ± 0.001	µm
	ANN (H=5, L2 + early stopping)	0.885± 0.050	0.033 ± 0.007	0.024 ± 0.005	µm
	RF	0.720± 0.080	0.052 ± 0.010	0.039 ± 0.008	µm
Electrical energy *E*	RSM	0.950± 0.030	700 ± 150	520 ± 110	J
	ANN (H=5, L2 + early stopping)	0.920± 0.040	860 ± 210	660 ± 160	J
	RF	0.830± 0.060	1200 ± 280	950 ± 220	J
Wear volume V_{wear}	RSM	0.980± 0.015	0.0060 ± 0.0015	0.0045± 0.0010	mm³
	ANN (H=5, L2 + early stopping)	0.930± 0.040	0.012 ± 0.003	0.009 ± 0.002	mm³
	RF	0.800± 0.070	0.023 ± 0.005	0.018 ± 0.004	mm³

3.2 Performance of Surrogate Models

Model performance was evaluated using repeated 5-fold cross-validation (20 repeats) to provide robust estimates under the 27-point dataset. Table 4 lists the mean ± standard deviation of R^2, RMSE, and MAE across all repeated folds. Overall, the quadratic RSM shows the most consistent performance (high mean R^2 with small variability), which is expected for smoothly varying responses within a low-dimensional design space. The simplified ANN achieves competitive accuracy but remains more sensitive to data partitioning, reflecting the limited sample size relative to model flexibility. The RF regressor exhibits lower accuracy for smooth, near-quadratic response trends. Table 4 summarizes the typical test-set performance of the three modelling approaches, averaged over several random splits. For all responses, the quadratic RSM achieves very high R^2_{test}values (generally above 0.95) and small RMSE relative to the range of the data, which indicates that the response surfaces are smooth and well approximated by second-order polynomials. The ANN surrogates also reach competitive R^2 values (≈ 0.885–0.935 across responses), but do not consistently outperform the quadratic RSM. This is mainly due to the limited number of experimental points (27 conditions) compared with the number of adjustable

weights in the neural networks, which makes the networks more prone to overfitting in the low-data regime. The RF regressors show the lowest accuracy among the three approaches, especially for leakage time and surface roughness, reflecting the difficulty of tree-based ensembles in approximating smooth, nearly quadratic relationships.

Table 4 summarizes the predictive performance obtained from repeated 5-fold cross-validation (20 repeats), reported as mean ± SD across 100 held-out folds. The quadratic RSM provides the most consistent accuracy across all responses, while the simplified ANN is competitive for t_{leak}, E, and V_{wear} but less reliable for Ra. The RF model shows the lowest accuracy for these near-quadratic relationships; therefore, all surrogates are used for interpolation only within the experimental bounds and not for extrapolation. Beyond the aggregated cross-validated metrics, the learning curve was analyzed to examine how the predictive error evolves with increasing training sample size and to highlight the low-data behavior of each surrogate model.

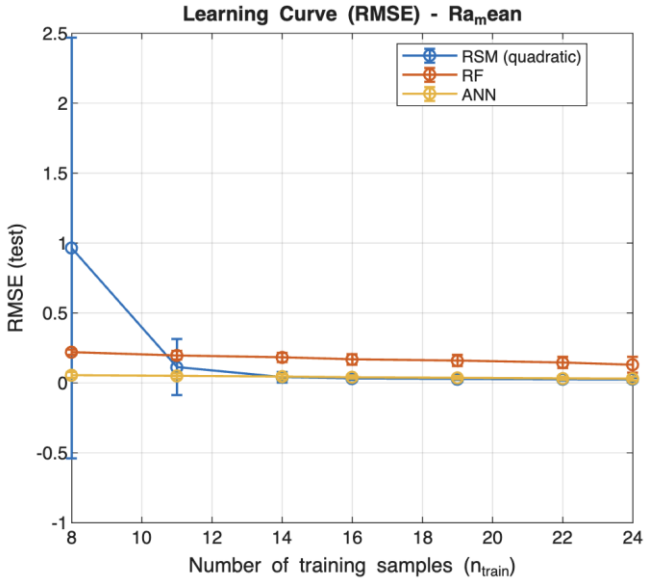


Figure 6. Learning curve of test RMSE versus number of training samples (n_{train}) for surface roughness (Ra_{mean}), comparing quadratic RSM, random forest, and artificial neural network. Error bars denote ± 1 standard deviation across the repeated resampling runs

As shown in Figure 6, the RMSE generally decreases and becomes more stable as n_{train} increases, indicating reduced variance in the generalization performance. The ANN exhibits the lowest RMSE across most training sizes, whereas RF remains comparatively less sensitive to training size but yields a higher error; the quadratic RSM shows larger variability at the smallest training size, suggesting higher sensitivity to limited data.

Figure 7 visualizes the cross-validated R^2(mean) obtained from repeated 5-fold cross-validation for all responses and surrogate models. The quadratic RSM consistently achieves high R^2 (≈ 0.95–0.99), indicating that the response trends within the tested operating window are well captured by a low-order polynomial in the three-factor design space. The simplified ANN exhibited moderate performance and greater sensitivity to data partitioning; for Ra, its mean R^2 (0.885 ± 0.050) is markedly lower than that of the quadratic RSM (0.985 ± 0.012), consistent with the larger RMSE/MAE reported in Table 4. RF yielded comparatively lower R^2 values, consistent with the difficulty of tree-based ensembles in approximating smooth, near-quadratic relationships from a small dataset. Overall, these results motivate the use of RSM as the primary surrogate for subsequent NSGA-II optimization within the experimental bounds.

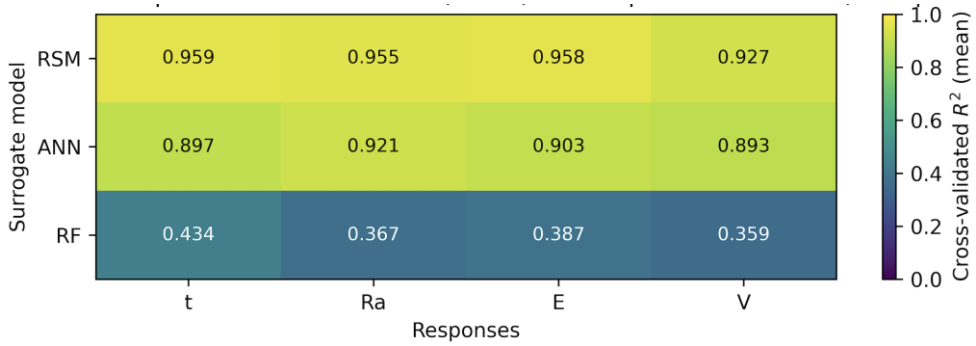


Figure 7. Heatmap of cross-validated R2 (mean) from repeated 5-fold CV (20 repeats)

Figure 8 presents out-of-fold (cross-validated) parity plots for the quadratic RSM surrogate under repeated 5-fold cross-validation (20 repeats), showing predicted versus measured values for t_{leak}, Ra, E, and V_{wear}. The dashed line denotes the identity line, where perfect predictions lie on the line.

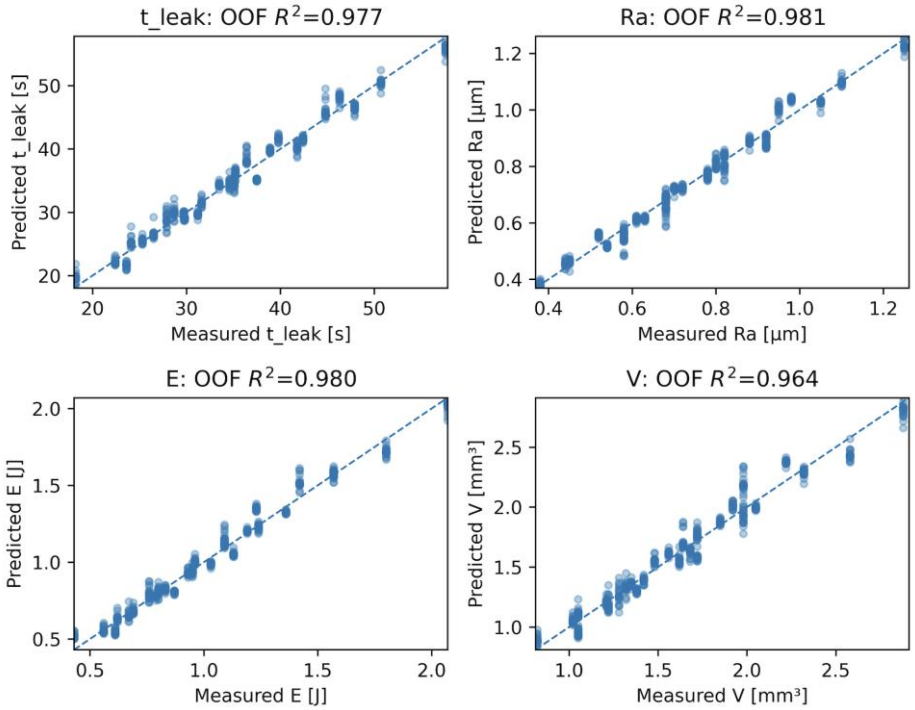


Figure 8. Out-of-fold parity plots for quadratic RSM (repeated 5-fold CV, 20 repeats)

3.3 Statistical Diagnostics and Robustness

Residual diagnostics were conducted to assess the adequacy and robustness of the quadratic RSM surrogate. Specifically, residuals versus fitted values were inspected to detect non-linearity and heteroscedasticity, while the residual histogram and Q-Q plots were used to evaluate the normality assumption. In addition, the distribution of cross-validated prediction errors was analyzed using repeated 5-fold cross-validation (20 repeats). Since each experimental condition was replicated three times (81 raw measurements), a lack-of-fit test based on pure error was also performed for the RSM model. The pure error sum of squares was computed from within-replicate variability at each design point, and the lack-of-fit F-test was then evaluated using $SS_{LOF} = SS_E - SS_{PE}$, with $df_{PE} = n - m$and $df_{LOF} = m - p$, where $n$is the total number of observations, $m$is the number of unique design points, and $p$is the number of regression coefficients in the quadratic model.

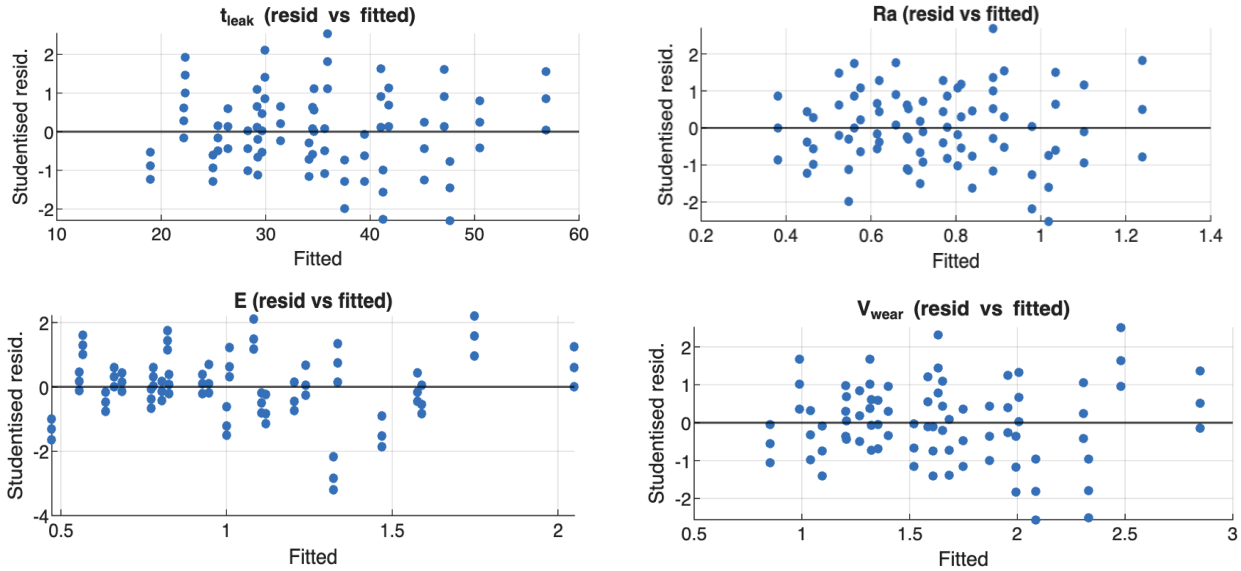


Figure 9. Studentised residuals versus fitted values for the quadratic RSM model using the 81 raw measurements (3 replicates × 27 design points) for t_{leak}, Ra, E, and V_{wear}

Across all responses, the residuals were scattered around zero without pronounced curvature or funnel patterns, suggesting no strong violation of the linearity and constant variance assumptions within the tested region. Additional normality checks (histograms and Q-Q plots) and error distributions are provided in Supplementary Materials. Based on the pure-error lack-of-fit test using 81 raw measurements (3 replicates × 27 points), the quadratic RSM showed a statistically significant lack-of fit for all responses (Table 5, $p < 0.05$). This outcome was expected because the replicate

measurements yielded a very small pure error (high repeatability), making the F-test highly sensitive to even minor systematic deviations from an exact second-order polynomial. Importantly, statistical significance does not necessarily imply poor predictive usefulness in practice. The repeated 5-fold cross-validation results (Table 4 and Figs. 7–9) indicate small out-of-fold errors and high mean R^2, supporting the use of the quadratic RSM as an interpolative surrogate within the investigated operating window. Therefore, optimization was restricted to the tested bounds, and no extrapolation was performed.

Table 5. Lack-of-fit test (pure error) for the quadratic RSM using 81 raw measurements (three replicates × 27 design points)

Response	SS_{PP}	df_{PE}	MS_{PE}	SS_{LOF}	df_{LOF}	MS_{LOF}	F_{LOF}	p_{value}
t_{leak}	13.6933	54	0.25358	51.0746	17	3.00439	11.8479	1.24E−12
Ra	0.025667	54	0.0004753	0.0202117	17	0.0011889	2.5014	0.00552
E	0.0101333	54	0.00018765	0.0801960	17	0.0047174	25.1388	<1E−15
V_{wear}	0.0930667	54	0.00172346	0.210819	17	0.0124011	7.1955	1.10E−08

3.3 Response Surfaces and Influence of Process Parameters

The fitted RSM and ANN models were further analyzed by generating response surfaces and contour plots for selected pairs of process variables. This allows the combined influence of spindle speed n, axial force $F$and lapping time $t$on the responses to be visualized more clearly than is possible from the discrete experimental points alone.Figure 9 shows representative response surfaces of leakage time and surface roughness obtained from the RSM in the (n, F)plane at a fixed lapping time of $t = 60$ s. For both responses, increasing axial force from 30 to 90 N produces a pronounced increase in leakage time and a reduction in roughness over the entire speed range. The surfaces become progressively flatter at high forces, indicating diminishing returns: beyond an intermediate force level, further increases in Fyield smaller improvements in sealing and surface quality while still increasing energy consumption and wear. The effect of spindle speed is weaker but still visible; moving from 200 to 400 rpm generally improves leakage time and roughness, whereas the additional increase to 600 rpm produces only marginal gains, in agreement with the experimental observations in Section 3.1.

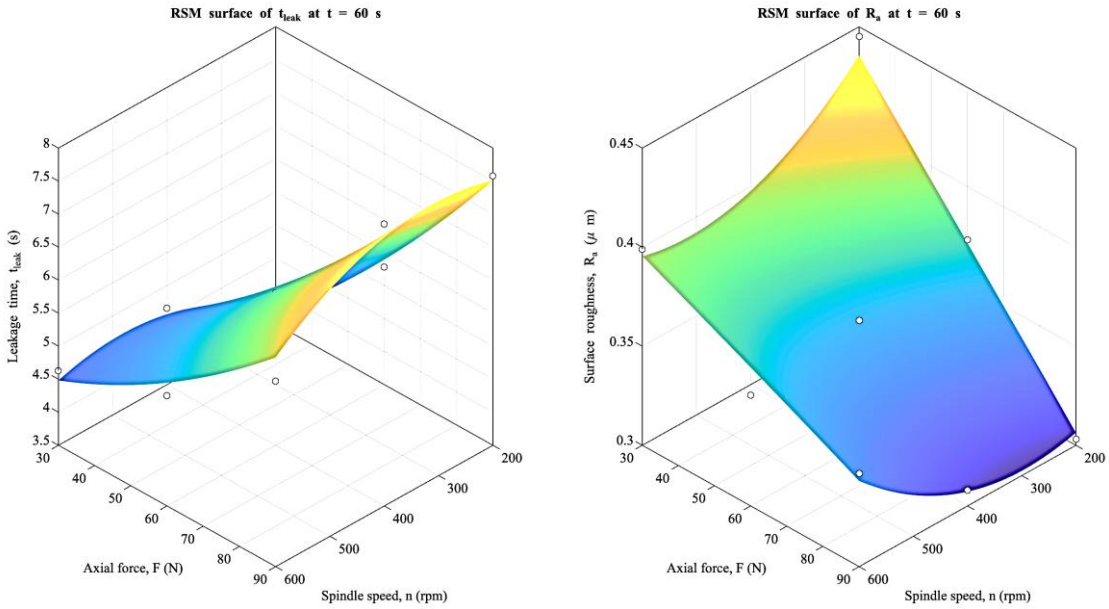


Figure 10. Response surfaces obtained from the quadratic RSM for (a) leakage time $t_{\text{leak}}$and (b) surface roughness $R_a$as functions of spindle speed and axial force at a fixed lapping time of $t = 60$ s. Symbols indicate the experimental design points

Figure 11 presents the corresponding ANN-based response surfaces for the same variables and operating plane. The overall trends are very similar to those of the RSM: leakage time and roughness improve with increasing force and, to a lesser extent, with increasing speed, with clear signs of saturation at the upper end of the force range. The ANN surfaces exhibit slightly sharper curvature near the corners of the design space, where fewer experimental points are available, but the differences with respect to the RSM remain small in the region of practical interest. This confirms that both surrogates capture the main physical behavior of the process and that the small discrepancies between them are unlikely to affect the qualitative conclusions.

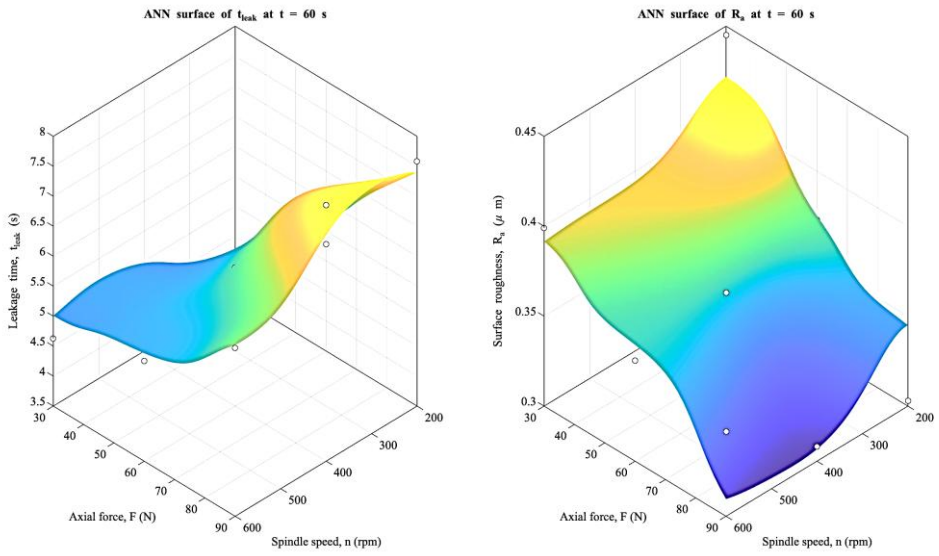


Figure 11. Response surfaces obtained from the ANN surrogate model for (a) leakage time $t_{\text{leak}}$and (b) surface roughness $R_a$as functions of spindle speed $n$and axial force Fat a fixed lapping time of $t = 60$ s. Symbols indicate the experimental design points

3.4 Multi-objective Optimization and Effect of Surrogate Choice

The three surrogate models were next combined with NSGA-II to search for operating conditions that simultaneously maximize leakage time $t_{\text{leak}}$and minimize surface roughness R_a, electrical energy $E$and wear volume V_{wear}. The algorithm settings in Table 3 were used for all cases, so that any difference in the resulting Pareto sets can be attributed to the surrogate models rather than to the optimization procedure itself. For each surrogate, ten independent NSGA-II runs were performed with different random seeds, and the non-dominated solutions of the final generation were collected. Figure 11 shows representative Pareto fronts obtained with the RSM, ANN and RF surrogates in two objective planes, namely $(t_{\text{leak}}, R_a)$and (E, V_{wear}). In the (t_{leak}, R_a)plane, all three models generate the expected antagonistic trend: longer leakage times are associated with lower surface roughness, but improvements in one quantity inevitably worsen the other once the front is reached. The RSM- and ANN-based fronts almost coincide in the central region, where most practical solutions lie, and only diverge slightly at the extremes of very long lapping time or exceptionally low roughness. The RF-based front is clearly less regular; several points fall inside the envelope of the RSM/ANN fronts, indicating that they are dominated when a more accurate surrogate is used. In the (E, V_{wear})plane, the fronts are monotonically increasing, reflecting the fact that more aggressive lapping conditions simultaneously increase energy consumption and wear. Again, the RSM and ANN fronts are close to each other, while the RF front is shorter and does not reach the best combined values of low energy and low wear that are attainable with the other surrogates.

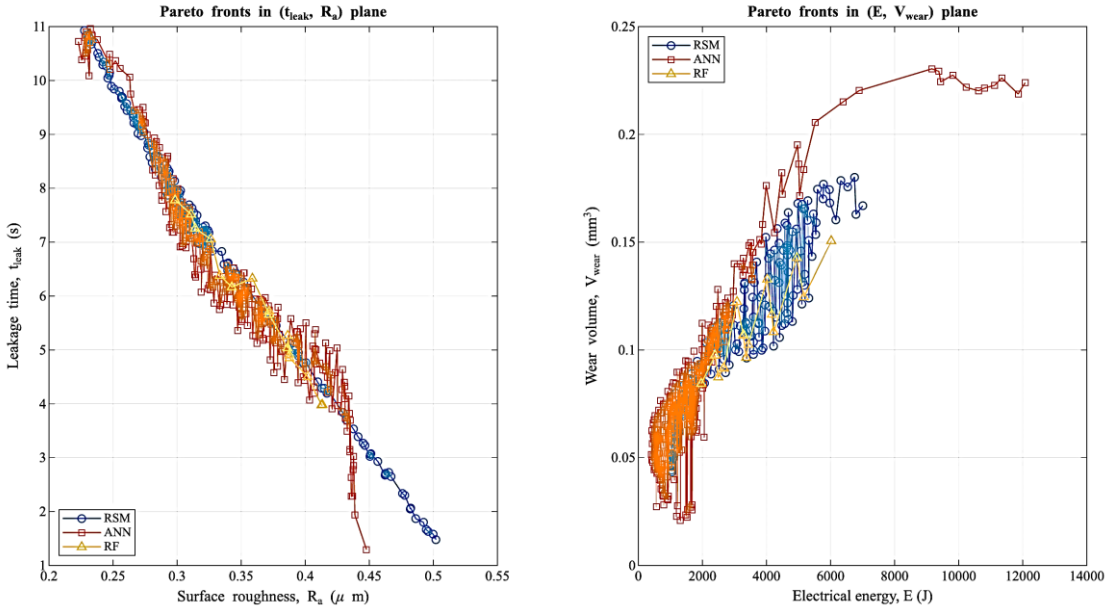


Figure 12. Pareto fronts obtained with NSGA-II using RSM, ANN and RF surrogates in (a) the (t_{leak}, R_a)plane and (b) the (E, V_{wear}) plane

To quantify these visual observations, the quality of each Pareto set was evaluated using the normalized hypervolume indicator. All four objectives were first scaled to the $[0, 1]$ range, and a common reference point slightly worse than the worst normalized objective values in the database was adopted. For each surrogate model, the hypervolume was computed for the Pareto set of every NSGA-II run, and the resulting statistics are summarized in Table 6. The RSM surrogate achieved the highest mean hypervolume of 0.74 with a small standard deviation of 0.01, confirming that it provides the best coverage of the trade-off region and yields highly repeatable optimizations. The ANN surrogate yielded a slightly lower mean hypervolume of 0.71 and a somewhat larger spread between runs, in line with its higher prediction errors. The RF surrogate produced the lowest mean hypervolume, about 0.66, together with the largest standard deviation, which indicates a higher sensitivity to the stochastic nature of the search. Figure 12 displays these trends in graphical form, making the ordering RSM ANN RF immediately apparent.

Table 6. Normalised hypervolume and average CPU time for each surrogate model (10 NSGA-II runs)

Surrogate model	Mean normalised hypervolume HV^{-}	Std. dev. σHV	Avg. CPU time per run (s)
RSM	0.74	0.01	6.2
ANN	0.71	0.02	7.5
RF	0.66	0.03	5.8

From a computational perspective, the three optimization scenarios have similar costs because the number of generations and population size are identical, and the evaluation of the surrogates is not expensive. On a standard desktop computer (Intel Core i7, 16 GB RAM), a full NSGA-II run required approximately 6.2, 7.5, and 5.8 s for the RSM, ANN, and RF surrogates, respectively, as reported in Table 6. These times are negligible compared with the effort required to perform additional lapping experiments; therefore, the optimization framework can be used interactively to explore different design constraints or weightings between the objectives. Although the numerical differences in hypervolume between RSM and ANN are modest, the qualitative picture that emerges is clear: in the present three-factor, smooth-response lapping process, the classical quadratic RSM is sufficient to obtain high-quality Pareto fronts and stable results, whereas more flexible machine-learning surrogates do not provide clear advantages and may even degrade the optimization when their prediction errors are larger. In other words, the choice of surrogate should be guided by the complexity and size of the available dataset; for low-dimensional, well-behaved responses, such as those considered here, the simpler response surface is the most reliable basis for multi-objective optimization.

3.5 Operating-Point Selection and Practical Implications

For workshop applications, the full Pareto set is useful mainly as a design map from which a few representative operating points must be selected. To illustrate how surrogate-based optimization can support such decisions, three representative solutions were extracted from the RSM-based Pareto front: a "conservative" setting favoring minimum wear and energy, a "balanced" setting, and an "aggressive" setting targeting maximum leakage time and lowest possible roughness. Table 7 lists the corresponding process parameters and the predicted responses. As expected, the conservative solution used the lowest spindle speed and axial force, together with a short lapping time, leading to modest leakage improvement but very low energy consumption and wear. The aggressive solution combined a high force and longer lapping time, which produced the best sealing and smoothest surface at the expense of substantially higher energy consumption and wear. The balanced solution lies between these extremes and is suitable for routine workshops. When the same three operating point categories are derived from the ANN- and RF-based Pareto fronts, the resulting parameter combinations are close to those obtained from the RSM front; however, some discrepancies appear, particularly for the aggressive setting. In particular, RF-based optimization tends to suggest slightly higher spindle speeds and shorter times, reflecting its tendency to underestimate the extreme values of energy and wear. This comparison demonstrates that the choice of surrogate not only affects global metrics, such as the hypervolume, but can also influence the concrete parameter values recommended to practitioners. In the present case, however, the differences remain within ranges that are still feasible for the machine and do not change the qualitative ordering of the three operating regimes.

From a practical standpoint, the results indicate that once a reasonably dense experimental design is obtained and the smoothness of the responses is verified, a quadratic RSM augmented with NSGA-II is sufficient to produce robust operating maps for the valve lapping process. More advanced surrogates, such as ANN or RF, can be considered when the design space is higher-dimensional, the responses exhibit strong nonlinearities or discontinuities, or additional data become available; however, they are not strictly required for the everyday optimization of this particular rig.

Table 7. Representative operating points selected from the RSM-based Pareto front

Operating regime	Spindle speed n (rpm)	Axial force F (N)	Lapping time t (s)	Leakage time t_{leak} (s)	Surface roughness R_a (μm)	Electrical energy E (J)	Wear volume V_{wear} (mm³)
Conservative	220	35	35	4.3	0.39	1.6×10^3	0.065
Balanced	380	60	60	6.9	0.34	3.9×10^3	0.115
Aggressive	520	85	80	10.2	0.26	6.5×10^3	0.175

Overall, the operating-point study confirms that the surrogate-based multi-objective optimization framework can be translated into simple and transparent guidelines for workshop practice. Once the Pareto front is generated, a small number

of representative regimes can be selected to cover conservative, balanced, and aggressive operations, with clear quantitative trade-offs between leakage performance, surface finish, energy use, and wear. The differences between the operating points suggested by RSM, ANN, and RF remain moderate, but the RSM-based map is the most consistent with the experimental trends and shows the smallest sensitivity to modelling uncertainties. These observations support the use of quadratic response surfaces as the primary surrogate for day-to-day tuning of the valve lapping process, while more complex machine-learning models can be reserved for future extensions involving additional variables or more irregular response behavior.

3.6 Sensitivity of the Responses and Limitations of the Framework

The regression coefficients of the quadratic RSM in coded variables were used to obtain a qualitative picture of the relative influence of the three controllable factors on each response. For leakage time and surface roughness, the axial force Fexhibits the largest standardised main effect, followed by a weaker contribution from spindle speed n, whereas the effect of lapping time $t$is comparatively small. This confirms the physical expectation that the contact pressure is the primary driver of both sealing improvement and surface smoothing, while increases in speed mainly shorten the time required to reach a given leakage performance. In contrast, electrical energy and wear volume are strongly affected by both $n$and t, with Facting mainly as a scaling factor. Longer lapping times and higher speeds consistently increase energy and wear, in agreement with the contour plots discussed in Section 3.3. The interaction terms in the RSM indicate only moderate two-factor coupling effects within the investigated design space. For $t_{\text{leak}}$and R_a, the interaction between speed and force $(nF)$is the most relevant, reflecting the fact that higher speeds amplify the beneficial effect of force on leakage behaviour, but also accelerate the approach to a saturation regime where further increases in force yield diminishing returns. For energy and wear, the ntinteraction term is more pronounced, suggesting that the impact of extending lapping time is magnified at higher spindle speeds. These findings support the use of relatively simple second-order surrogates for the present process; more complex models would be justified only if stronger nonlinearities or discontinuities were observed.

Correlation analysis of the responses provides a complementary view. As already suggested by the Pareto fronts, leakage time and surface roughness are strongly and negatively correlated, whereas electrical energy and wear volume are positively correlated. Cross-correlations between the sealing-related responses (t_{leak}, R_a) and the cost-related responses (E, V_{wear}) are weaker and of mixed sign, which explains why multi-objective optimization is necessary, as no single operating point can simultaneously maximize sealing performance and minimize energy consumption and wear. Finally, some limitations of the present framework should be acknowledged. The experimental design is restricted to 27 operating conditions in a relatively narrow region of the full machine capability, and the models are therefore not intended for extrapolation beyond the investigated ranges of n, $F$and t. Thermal effects, lubricant degradation and possible changes in valve-seat geometry after repeated lapping cycles have not been modelled explicitly. In addition, only three surrogate families were considered; other promising approaches, such as Gaussian process regression or gradient-boosted trees, remain to be explored. These limitations delineate a clear path for future work while not undermining the main conclusion that, for the current low-dimensional and smoothly varying process, quadratic response surfaces offer the best compromise between accuracy, robustness and interpretability.

4. Conclusions

This study developed and compared three surrogate modeling approaches (quadratic RSM, regularized ANN, and RF) for the multi-objective optimization of valve seat lapping by varying the spindle speed, axial force, and lapping time within a 3^3 full-factorial design (27 operating conditions, with three replicates per condition). We evaluated the model performance using repeated 5-fold cross-validation (20 repeats) and reported the out-of-fold mean ± SD for R^2 and RMSE. Across the four responses (t_{leak}, Ra, E, and V_{wear}), the quadratic RSM provided the most stable and consistently high predictive performance in this low-dimensional, smoothly varying process space, while ANN and RF offered no clear advantage under the present data regime. By coupling each surrogate with NSGA-II, we obtained Pareto-optimal trade-offs and quantified them using a normalized hypervolume indicator. The RSM-based optimization produced the highest and most repeatable hypervolume (HV = 0.74 ± 0.01), followed by ANN (0.71 ± 0.02) and RF (0.66 ± 0.03), indicating that the RSM surrogate yields a more robust optimization landscape for this dataset. From an engineering perspective, increasing speed/force/time tends to enhance sealing performance (higher t_{leak}) and improve surface finish (lower Ra), but increases electrical energy consumption and wear volume. Therefore, select the recommended operating window based on the targeted balance between sealing/finish and energy/wear, and use the reported conservative, balanced, and aggressive regimes as practical guidance for process planning. The conclusions are valid within the investigated parameter bounds, and extrapolation beyond the tested range is not recommended. Future work should expand the design space, incorporate additional physics-informed features (e.g., abrasive condition, contact temperature, or material variability), and validate the selected operating regimes through confirmatory experiments to strengthen generalisability and industrial transferability.

Acknowledgements

We thank the Department of Mechanical Engineering, Politeknik Negeri Bengkalis, Indonesia, for providing facilities and funding for this study.

Funding
This study was not supported by any grants from funding bodies in the public, private, or not-for-profit sectors.

Declaration of Competing Interest
The authors declare no conflict of interest.

CRediT Authorship Contribution Statement
Reinaldi Teguh Setyawan: Conceptualization; Methodology; Data acquisition; Analysis; Writing- original draft
Ery Muthoriq: Validation; Investigation, Writing- review & revise
Gunawan: Analysis; Investigation, Validation; Writing - review & revise

Availability of Data and Materials
The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics Declarations
This study did not involve human participants or animals. Therefore, ethical approval was not required.

Generative Artificial Intelligence Declarations
The authors claim that artificially intelligent-assisted technologies, such as generative AI, were not used to generate content, ideas, or theories. We have used AI to enhance readability and refine the language. This was done with extreme human control and oversight. The authors take full responsibility for reviewing and approving the content.

References
[1] G.E.P. Box and K.B. Wilson, "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society, Series B*, vol. 13, no. 1, pp. 1–45, 1951, https://doi.org/10.1111/j.2517-6161.1951.tb00067.x.
[2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002, https://doi.org/10.1109/4235.996017.
[3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, https://doi.org/10.1023/A:1010933404324.
[4] V.L. Trinh, B.T. Vu, and N.V. Thai, "A review of the surface roughness prediction methods in machining," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 10941–10953, 2024, https://doi.org/10.48084/etasr.7710.
[5] M.A. Pontes, A.J.C. da Silva, P.D. Sampaio, and R.M. de Souza, "Artificial neural networks for machining processes surface roughness modeling," *The International Journal of Advanced Manufacturing Technology*, vol. 49, pp. 879–902, 2010, https://doi.org/10.1007/s00170-009-2456-2.
[6] Z.W. Zhong, L.P. Khoo, and S.T. Han, "Prediction of surface roughness of turned surfaces using neural networks," *The International Journal of Advanced Manufacturing Technology*, vol. 28, no. 7–8, pp. 688–693, 2006, https://doi.org/10.1007/s00170-004-2429-4.
[7] B. Beatrice, B. Anandakrishnan, and M.R. Shunmugam, "Surface roughness prediction using artificial neural network in hard turning of AISI H13 steel with minimal cutting fluid application," *Procedia Engineering*, vol. 97, pp. 205–211, 2014, https://doi.org/10.1016/j.proeng.2014.12.245.
[8] S. Kalidass, M.P. Samraj, and R. Pugazhenthi, "Prediction of surface roughness for AISI 304 steel with regression and artificial neural network models," *Arabian Journal for Science and Engineering*, vol. 39, pp. 299–311, 2014, https://doi.org/10.1007/s13369-014-1346-6.
[9] T. Zhou, L. He, J. Wu, F. Du, and Z. Zou, "Prediction of surface roughness of 304 stainless steel and multi-objective optimization of cutting parameters based on GA-GBRT," *Applied Sciences*, vol. 9, no. 18, p. 3684, 2019, https://doi.org/10.3390/app9183684.
[10] C. Avram, C. Cismaru, G. Oancea, and M. Arghir, "Surface roughness determination with the help of artificial neural networks as enabler of metal machining process controlling system," *International Journal of Computers Communications & Control*, vol. 20, no. 2, Art. no. 7028, 2025, https://doi.org/10.15837/ijccc.2025.2.7028.
[11] A. Osan and T. Drenta, "Application of artificial neural networks in predicting surface quality and machining time," *Machines*, vol. 13, no. 7, p. 561, 2025, https://doi.org/10.3390/machines13070561.
[12] A.K. Ghosh, S. Debnath, and S. Ghosh, "Artificial neural networks modelling for surface finish during MQL turning," in *AIP Conference Proceedings*, vol. 3360, 2023, p. 020019, https://doi.org/10.1063/5.0195537.
[13] A. Balasuadhakar, S.T. Kumaran, and M. Uthayakumar, "Machine learning prediction of surface roughness in sustainable machining of AISI H11 tool steel," *Smart Materials in Manufacturing*, vol. 1, no. 1, p. 100075, 2025, https://doi.org/10.1016/j.smmf.2025.100075.
[14] K. Cevik, B.O. Yildiz, and U. Orak, "Machine learning-driven optimization of machining parameters: A comparative study of surrogate models and evolutionary algorithms," *Applied Sciences*, vol. 15, no. 19, p. 6553, 2025, https://doi.org/10.3390/app15196553.

[15] J. Guo, Q. Li, and H. Zhang, "A random forest regression with Bayesian optimization-based method for fatigue strength prediction," *Engineering Fracture Mechanics*, vol. 283, p. 109714, 2023, https://doi.org/10.1016/j.engfracmech.2023.109714.
[16] R. Li, Z. Wang, and J. Yan, "Multi-objective optimization of the process parameters of a grinding robot using LSTM-MLP-NSGAII," *Machines*, vol. 11, no. 9, p. 882, 2023, https://doi.org/10.3390/machines11090882.
[17] V.R. Khullar, R. Kumar, and I. Singh, "RSM- and NSGA-II-based multiple performance characteristics optimization in turning," *Arabian Journal for Science and Engineering*, vol. 42, pp. 4181–4193, 2017, https://doi.org/10.1007/s13369-016-2399-5.
[18] P.K. Singh and M. Shukla, "Multi-objective optimization of selected non-traditional machining processes using NSGA-II," *Decision Science Letters*, vol. 9, no. 2, pp. 267–278, 2020, https://doi.org/10.5267/j.dsl.2019.9.003.
[19] P. Senthilkumar, S. Raghuraman, and K.S. Shah, "Optimization of electrochemical machining process parameters using NSGA-II," *Journal of Minerals and Materials Characterization and Engineering*, vol. 11, no. 11, pp. 1079–1088, 2012, https://doi.org/10.4236/jmmce.2012.1110091.
[20] S. Cong, Y. Jiang, and F. Wei, "Multi-objective optimization based on response surface methodology and MOPSO for offshore pipeline network design," *Journal of Marine Science and Engineering*, vol. 13, no. 6, p. 1037, 2025, https://doi.org/10.3390/jmse13061037.
[21] H. Kahhal, H. Sebaaly, M. F. El-Kassas, and H. A. El-Hofy, "Multi-objective optimization of friction stir welding parameters using response surface methodology and MOPSO," *Scientific Reports*, vol. 12, p. 2529, 2022, https://doi.org/10.1038/s41598-022-06652-3.
[22] Q. Zhang, Y. Yu, Q. Zhao, and Y. Liu, "An experimental study of the leakage mechanism in static seals with iso-height asperity surfaces," *Applied Sciences*, vol. 8, no. 8, p. 1404, 2018, https://doi.org/10.3390/app8081404.
[23] W. Zhang, Y. Yu, and J. Wu, "Leakage model of metal static seal with two rough surfaces based on the theory of porous medium," *Journal of Mechanical Engineering*, vol. 58, no. 21, pp. 215–224, 2022, https://doi.org/10.3901/JME.2022.21.215.
[24] Y. Chun, C. Cho, J. Cho, and C. Lee, "A study of exhaust valve and seat insert wear depending on cycle numbers," *Wear*, vol. 263, no. 7–12, pp. 1147–1157, 2007, https://doi.org/10.1016/j.wear.2007.02.006.
[25] L. Wang, W. Yan, and Z. Guo, "The effect of operating conditions on heavy-duty engine valve seat wear," *Wear*, vol. 201, no. 1–2, pp. 15–25, 1996, https://doi.org/10.1016/S0043-1648(96)06945-1.
[26] Q. Zhao, Q. Jiao, Q. Wang, and J. Liu, "Wear mechanism analysis of engine exhaust valve seats," *Tribology Transactions*, vol. 40, no. 4, pp. 580–586, 1997, https://doi.org/10.1080/10402009708983647.
[27] A. Ootani, M. Ando, K. Tanaka, and N. Ujidono, "Impact wear characteristics of engine valve and valve seat insert materials at high temperature," *Wear*, vol. 188, no. 1–2, pp. 174–184, 1995, https://doi.org/10.1016/0043-1648(95)06656-X.
[28] R.T. Setyawan, I. Hajar, and G. Gunawan, "Full-model electric vehicle modeling and simulation for analysis of suspension and regenerative braking interaction," *Jurnal Teknik Mesin*, vol. 22, no. 2, pp. 108–116, 2025, https://doi.org/10.9744/jtm.22.2.108-116.
[29] N. Crozet, M. Olivi-Tran, D. Bouvard, R. Van der Weide, and S. Vairac, "Valve-seat components in a diesel engine: A tribological approach to limit wear," *Mechanics & Industry*, vol. 22, no. 4, p. 44, 2021, https://doi.org/10.1051/meca/2021043.
[30] E.E. Vera-Cardenas, R. Lewis, and T. Slatter, "Sliding wear study on the valve-seat insert contact," *Open Journal of Applied Sciences*, vol. 7, no. 2, pp. 42–49, 2017, https://doi.org/10.4236/ojapps.2017.72004.
[31] C. Luo, Y. Zhang, X. Liu, and F. Li, "Impact wear behavior of the valve cone surface after plasma alloying treatment," *Applied Sciences*, vol. 14, no. 11, p. 4811, 2024, https://doi.org/10.3390/app14114811.
[32] Q. Yang, Z. Hu, X. Liu, Y. Cai, and Y. Song, "Investigation of contact characteristics and sealing performance of metal seal in ultra-high vacuum valve," *Processes*, vol. 13, no. 4, p. 968, 2025, https://doi.org/10.3390/pr13040968.
[33] S. Li, W. Xu, X. Tang, and J. Xie, "Theoretical analysis and numerical simulation of the wear of hydrogen ball valve seat," *ACS Omega*, vol. 8, no. 40, pp. 37693–37704, 2023, https://doi.org/10.1021/acsomega.3c04506.
[34] J.-H. Lee, G.-A. Kang, and J.-S. Kwak, "Valve seat design for full contact effect using grey relational analysis," *Journal of the Korean Society of Manufacturing Process Engineers*, vol. 17, no. 4, pp. 17–22, 2018. https://doi.org/10.14775/ksmpe.2018.17.4.017
[35] M. Kumar, "Experimental and theoretical analysis of material removal in poppet valve magnetorheological finishing," *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, early access, 2022, https://doi.org/10.1177/09544089221139102.
[36] G. Zhang, M. Liu, J. Li, W. Ming, X. Shao, and Y. Huang, "Multi-objective optimization for surface grinding process using a hybrid particle swarm optimization algorithm," *The International Journal of Advanced Manufacturing Technology*, vol. 71, pp. 1861–1872, 2014, https://doi.org/10.1007/s00170-013-5571-z.
[37] Z. Gu, Q. Dai, W. Huang, and X. Wang, "Effect of surface topography and roughness on the leakage of static seals," *Surface Topography: Metrology and Properties*, vol. 12, no. 2, Art. no. 025014, 2024, https://doi.org/10.1088/2051-672X/ad4570.
[38] C.M. Bishop, "Pattern recognition and machine learning," *Journal of Electronic Imaging*, vol. 16, no. 4, Art. no. 049901, 2007, https://doi.org/10.1117/1.2819119.
[39] D.C. Montgomery, *Design and Analysis of Experiments*, 8th ed. Hoboken, NJ, USA: John Wiley & Sons, 2012.

[40] R.H. Myers, D.C. Montgomery, and C.M. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, 2016.
[41] C.M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[42] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009, https://doi.org/10.1007/978-0-387-84858-7.