The Galileo Ferraris Contest: A Benchmark Initiative for Data-Driven Multi-Physics Modeling of Traction Electric Motors

Luigi Solimene[1], Simone Ferrari[1], Riccardo Torchio[3], Costanza Anerdi[2], Fabio Freschi[1], Luca Giaccone[1], Gianmarco Lorenti[1], Francesco Lucchini[3], Spyridon Mallios[4], Patrick Lombard[4], Paolo Di Barba[5], Arash Ghafoorinejad[5], Maria Evelina Mognaschi[5], Dimitrios Loukrezis[6], Eric Diehl[7,9], Eniz Museljic[8], Aylar Partovizadeh[7], Manfred Kaltenbacher[8], Sebastian Schöps[7], Soo-Hwan Park[10], Chan-Woo Kim[10], Ji-Hyeon Lee[11], Younjae Hwang[11], Min-Ro Park[12], Myung-Seop Lim[11], Sina Toghranegar[13], Ruth V. Sabariego[13], Geert Deconinck[13], Hussain Kazmi[13], Zhi Gong[14], Zuqi Tang[14], Abdelkader Benabou[14], Guillaume Verez[15], Tohid Sharifi[16], Ali Jamali-Fard[16], Deniz Bilgili[17], Jasmin Jelovica[17], Paolo Manfredi[18], Riccardo Trinchero[18], Amir Akbari[19], Shahin Mahdiyounrad[19], David A. Lowther[19], Siyuan Sun[20], Toshiaki Koike-Akino[21], Ye Wang[21], Tatsuya Yamamoto[22], Yusuke Sakamoto[22], Bingnan Wang[21], Ankit Kumar Singh[23], Sagnik Dey[23], Swayambhu Mohanty[23], Aditya Anil Pulikottil[23], Asmijit Saha[23], Ankit Dalal[23], Yuki Sato[24], Hidenori Sasaki[25], Shingo Hiruma[26], Akito Maruo[27], Elena Blazhevska Ivanoski[28], Eymen Ipek[28], Niels Koester[28], Shahryar Rahnamayan[29], Piergiorgio Alotto[3], Gianmario Pellegrino[1], Maurizio Repetto[1]

See the Acknowledgments section for the full list of author affiliations.

The Galileo Ferraris Contest is a benchmark initiative aimed at evaluating data-driven surrogate modeling methodologies for multi-physics simulation of traction electric motors. The design of such motors traditionally relies on high-fidelity finite-element simulations, which are accurate but introduce severe computational bottlenecks for large-scale design exploration. The use of emerging surrogate modeling strategies offers a promising path to overcome these limitations. Yet, a systematic and fair comparison of their capabilities for realistic electric motor design is still lacking. Built upon an open-access dataset including electromagnetic, thermal, and structural results for three families of V-shaped interior permanent magnet motors, the contest provided a standardized testbed to assess interpolation, extrapolation, and innovation capabilities of surrogate models. A uniform multi-objective optimization and FEM validation pipeline was applied to all participant models, ensuring fair comparison across different machine learning strategies. The results show that different approaches successfully reproduced the input–output relationships of the reference motors, achieving accurate predictions even in unseen design regions. Moreover, when integrated into the optimization loop, surrogate models identified Pareto-optimal configurations beyond the resolution of the original dataset, enabling physically-consistent design exploration at negligible computational cost compared to direct finite element analysis. By establishing a reproducible framework, the Galileo Ferraris Contest preliminarily validated surrogate models as essential tools for electric motor optimization and advanced the standardization of data-driven multi-physics design workflows for future research.

***Index Terms*—Surrogate modeling, Data-driven design, Multi-physics simulation, Interior permanent magnet motor (IPM), Optimization benchmark, Interpolation, Extrapolation, Design innovation, Machine learning.**

I. Introduction

The design of traction electric motors is becoming increasingly challenging due to the simultaneous need for high performance, compactness, and multi-objective trade-offs across physical domains. In this context, electric motor optimization requires, at least, the integration of electromagnetic, thermal, and mechanical considerations, which are often in conflict. Traditional design workflows based on repeated finite element method (FEM) simulations can be accurate but are typically computationally expensive. This becomes particularly limiting when exploring large design spaces or running iterative optimization processes. As a result, surrogate models trained on simulation data are gaining attention as efficient alternatives that can accelerate the design loop, enabling fast predictions of performance indicators under varying design conditions. Several initiatives are sprouting, trying to heed the need for sensible and accurate data. Hosting and availability of this process must be provided through platforms such as IEEE DataPort [1] or Zenodo [2].

In recent years, numerous data-driven modeling techniques have been proposed for electrical machines, adopting both black-box regression methods and physics-informed approaches [3]–[5]. However, the lack of standardized datasets and public benchmarks has hampered the comparison and reproducibility of these methods.

To address this gap, the *Galileo Ferraris Contest* (GalFer Contest) [6], named after the Italian engineer *Galileo Ferraris* [7], a pioneer of alternating-current electric motors, was launched as a community-driven initiative to benchmark surrogate modeling methodologies for traction Interior Permanent

Corresponding author: Luigi Solimene (e-mail: luigi.solimene@polito.it).
This work has been submitted to IEEE Access for possible publication. This preprint has not undergone peer review.

Magnet (IPM) motors. Initially based on the *International Compumag Society* community [8], the interest around the initiative has spread around in the computational and machine learning researchers.

The GalFer Contest is built upon the open-access dataset generation procedure presented in [9], which provides thousands of samples combining electromagnetic, thermal, and structural simulation results, and it represents a standard testbed for evaluating the generalization, interpolation, and innovation capabilities of data-driven models.

To focus attention on these aspects in the research community, a contest was created and some industrial sponsorships have been collected allowing to set up some money prizes for best performance of *surrogate* or *data driven* models. Participants were asked to submit surrogate modeling procedures capable of predicting seven performance metrics, including torque, torque ripple, temperature, stress, and active material usage, from a compact set of geometric and electrical input variables. The GalFer Contest was structured into three tracks:

- Interpolation, evaluating model accuracy within the domain of a single motor family;
- Extrapolation, testing the model's ability to generalize to unseen motor configurations;
- Novelty, assessing methodological originality and usefulness for design insight.

To ensure rigorous and fair evaluation, a standardized multi-objective optimization and FEM validation pipeline was applied to all models, based on the approach proposed in [10].

The GalFer Contest was inspired in spirit by initiatives such as the MagNet Challenge [11], but with a specific focus on multi-physics modeling for motor design. By encouraging open data, reproducibility, and methodological comparison, the GalFer Contest aims to serve as a reference benchmark for future developments in the field.

This contribution has been thought, after the GalFer Contest conclusion, with the aim of dissemination of its main results. The paper has this structure: after a brief description of the dataset and workflow generation procedure, in Section II, the Section III details the GalFer Contest organization, evaluation criteria, and ranking procedure. Sections IV and V summarize the contributions and performance of the participating teams. Eventually, Section VI discusses the key insights from the GalFer Contest while Section VII draws some preliminary conclusions on the work and outlines possible directions for future editions.

II. Dataset and Workflow

The dataset comprises three motor families, Motor A, Motor B, and Motor C, each representative of a different application scenario within the electric mobility domain:

- **Motor A**: traction motor of full electric powertrain, with high torque and power ratings.
- **Motor B**: moto-generator of full hybrid powertrain with reduced size and moderate performance targets.
- **Motor C**: moto-generator of mild hybrid powertrain, characterized by compact geometry and lower nominal values.

All three motors share the same V-shaped IPM topology but differ in key design parameters such as stator outer diameter, stack length, pole pair number, and slot/pole/phase configuration. Table I summarizes their main characteristics.

The datasets used in the GalFer Contest were constructed by sampling a parametric design space defined by eight input variables, covering both geometric and electrical characteristics of a reference IPM motor with V-shaped rotor topology [9]. The input parameters and their ranges are reported in Table II. They include rotor and stator dimensions as well as the phase angle of the supply current.

To ensure a uniform and space-filling sampling of the design space, a Sobol' sequence was adopted. This quasi-random sampling technique generates low-discrepancy sequences that distribute points evenly across high-dimensional domains, providing improved coverage compared to regular grids or purely random methods [12].

For each sampled design, a multiphysics finite-element workflow was used to simulate the electromagnetic, thermal, and structural behavior. The workflow is managed in the open-source environment SyR-e [13], which integrates various tools for the analyses, such as FEMM [14] for electromagnetic analysis and the MATLAB PDE Toolbox [15] for structural analysis. Full details on the dataset construction, model assumptions, and simulation setup can be found in [9].

The overall computational cost of the FEM campaign for a 4096 configuration dataset was approximately 122 hours of computation time and was performed on a workstation equipped with an Intel Xeon Cascade Lake processor (32 logical cores) and 128 GB of RAM.

Dataset availability

The datasets on the three motors used in the GalFer Contest are publicly available on Zenodo [16] under the license CC BY 4.0.

III. Contest Structure and Evaluation Methodology

The GalFer Contest was designed to benchmark data-driven surrogate modeling methodologies in a realistic multi-physics design context. The competition focused on three distinct tracks: *Interpolation*, *Extrapolation*, and *Novelty*. These tracks targeted different aspects of model performance and innovation and were supported by a rigorous evaluation framework.

A. Tracks Overview

- Interpolation: Assessed the ability of a surrogate model to capture the input–output relationship of a single motor family dataset (Motor A), through a multi-objective optimization procedure.
- Extrapolation: Tested the model's ability to generalize across different motor families by training on Motors A and B and predicting on Motor C, a target almost unseen during training.
- Novelty: A track evaluated by an international advisory board based on the originality, insightfulness, and

TABLE I
RATED SPECIFICATIONS AND FIXED PARAMETERS OF THE MOTOR FAMILIES.

Parameter	**Motor A**	**Motor B**	**Motor C**
Application	Full Electric	Full Hybrid	Mild Hybrid
Rated Torque [Nm]	236	87	165
Rated Power [kW]	120	36.9	42
Max Speed [rpm]	15000	13500	12000
Stator Outer Diameter [mm]	225	264	291.3
Stack Length [mm]	134	50	61.7
Pole Pairs	3	4	4
Slots/Pole/Phase	3	2	2

TABLE II
INPUT DESIGN VARIABLE RANGES FOR MOTOR FAMILIES A, B, AND C.

Design Variable	**Motor A**	**Motor B**	**Motor C**
Barrier position d_α [p.u.]	[0.65, 0.85]	[0.40, 0.85]	[0.50, 0.60]
Barrier width h_c [p.u.]	[0.3, 0.7]	[0.3, 0.7]	[0.3, 0.4]
Rotor radius r [mm]	[60, 78]	[72, 90]	[78, 109]
Tooth width w_t [mm]	[3.8, 6.3]	[5.0, 9.0]	[3.4, 10.0]
Tooth length l_t [mm]	[15.0, 22.5]	fixed 12.36	[15.2, 32.0]
Slot opening w_o [p.u.]	[0.1, 0.4]	[0.1, 0.4]	[0.1, 0.4]
Barrier shift $dxIB$ [mm]	[–4, 6]	[–4, 8]	[–3, 6]
Current phase angle γ [°]	[30, 60]	[35, 75]	[35, 75]

efficiency of each team's approach. Being based on a *qualitative* assessment, this track is outlined but its results are not presented in this paper.

B. Contest Problem Definition

As summarized in Section II (see Tables I and II), the dataset provided to participants included simulation results for three different IPM traction motor families: Motor A, Motor B, and Motor C. Each motor configuration was characterized by a set of 8 input variables and 7 output performance metrics obtained via FEM-based multi-physics simulations [9]. For Motor families A and B, each dataset included 4096 configurations covering a broad range of design parameters. For Motor family C, only 256 configurations were shared, creating a low-data regime that challenged the generalization capability of surrogate models in the Extrapolation track. The outputs were:

1) Average torque (T)
2) Torque ripple (T_r)
3) Mass of copper (m_Cu)
4) Mass of magnets (m_mag)
5) Power factor ($\cos\varphi$)
6) Von Mises equivalent stress (VM)
7) Maximum winding temperature ($Temp$)

The design problem was formulated as a seven-objective optimization, with the goal of maximizing torque and power factor, while minimizing the other five objectives. Realistic constraints were enforced:

- $VM \leq 450\,\mathrm{MPa}$
- $Temp \leq 180\,^\circ\mathrm{C}$

The datasets were the only source of information for building and training the surrogate models: no external data or additional FEM models or simulations were provided to the participants.

C. Quantitative Evaluation Procedure

The evaluation procedure is based on a multi criteria methodology that is the same for the two tracks: Interpolation and Extrapolation mainly based on the approach proposed in [10].

1) Interpolation

In the Interpolation track, each team was asked to submit a data-driven surrogate model capable of emulating the input–output mapping for Motor family A. Teams were allowed to submit their procedure in either Python or MATLAB. To ensure a fair and uniform evaluation, all procedures were ultimately executed within a common NSGA-III optimization framework implemented in PlatEMO (MATLAB) [17]. The evaluation pipeline, outline in Figure 1, was the following:

1) Model submission: Each k^th team provided a single callable procedure (MATLAB .m function or Python .py script) able to accept an input vector composed of the 8 geometric and supply variables of Motor family A listed in Table II and 4 additional characteristic motor parameters describing the overall dimensions and magnetic properties of the motor listed in Table I, for a total of 12 scalar inputs, and to return the 7 output performance metrics defined in the dataset.
2) Integration with NSGA-III: All models were linked to PlatEMO NSGA-III implementation [17], using a wrapper that allowed calling Python functions from MATLAB (for Python models). The design space was defined according to the input variables of Motor family A.
3) Optimization runs: For each team, the NSGA-III optimizer was executed $R = 10$ times. All runs produced a set of non-dominated candidate designs $NDx_\mathrm{s}^{(\mathrm{k})}$.
4) FEM validation: The candidate solutions from all teams were validated using the original FEM-based analysis

workflow. Configurations violating physical constraints $(VM, Temp)$ after FEM validation were discarded.

5) Pareto front construction: From the validated points Y_{FEM}, the Best-So-Far Pareto Front (PF_{bsf}) was constructed, representing the ground-truth optimal front.

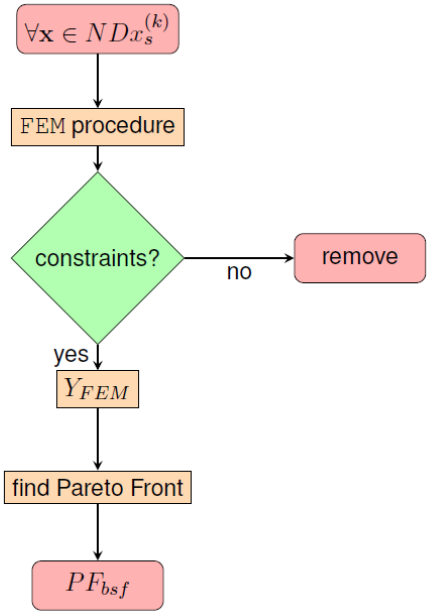

Fig. 1. Flowchart of the steps of the evaluation procedure.

Once the PF_{bsf} has been computed, the quality of each team contributions to it can be measured, following the approach proposed in [10]. Each team was scored using:

- Coverage: fraction of PF_{bsf} points generated by the team.
- Inverse Generational Distance (IGD): average distance between PF_{bsf} and the team's PF.
- Hypervolume (HV): volume dominated by the team's PF in objective space.

Finally, teams were ranked using a Pareto-based approach in metric space.

2) Extrapolation

The Extrapolation track assessed each model's ability to generalize across motor families. Participants received complete datasets (4096 configurations) for Motor families A and B, and a reduced dataset (256 configurations) for Motor family C, which the team used to retrain their models.

Each team submitted a surrogate model capable of operating in a 12-dimensional input space: beyond the 8 geometric and supply variables shown in Table II, 4 characteristic parameters describing the motor overall dimensions and magnetic parameters, contained in Table I, were added. On the basis of the inputs, the same 7 output quantities as in the Interpolation track had to be estimated. The models were again executed within the PlatEMO NSGA-III framework using the same MATLAB-based infrastructure.

The evaluation procedure was identical to the Interpolation case: Optimization, FEM validation, construction of the Best-So-Far Pareto Front (PFbsf), computation of Coverage, IGD, and Hypervolume, and final ranking via Pareto fronts and positional score.

This setup allowed for a consistent comparison of model performance in both interpolation and extrapolation regimes.

D. Novelty Assessment

The Novelty track aimed to recognize original and insightful contributions in the field of data-driven modeling for electrical machine design. Unlike the Interpolation and Extrapolation tracks, which were quantitatively evaluated, this track was assessed qualitatively based on written reports. Each participating team submitted a technical document (max 5 pages, IEEE double-column format) describing:

- The methodology and tools used.
- The architecture and training of the surrogate model.
- The innovative aspects of their approach.

The reports were independently evaluated by the GalFer Contest Advisory Board, composed of international experts in electromagnetics, machine learning, and electric motors and drives. The following criteria were considered:

1) Novelty of the data-driven method or its application to the motor design domain.
2) Design insight: the ability of the approach to support or enhance design decision-making beyond traditional methods
3) Efficiency: smart use of the provided training data, with respect to performance, complexity, or generalization.

Each criterion was scored individually by the advisory board. A three-dimensional metric space was then defined using these scores, and teams were ranked by applying the same Pareto-based and average positional scoring procedure used in the other tracks. This ensured a consistent and fair selection of the most promising contributions.

E. Performance Metrics and Ranking

The final rankings for the Interpolation and Extrapolation tracks were determined through a common multi-criteria procedure based on three metrics computed for each team:

- m_1 – Coverage: The number of points in the Best-So-Far Pareto Front (PF_{bsf}) contributed by the team:

$$m_1^{(k)} = N^{(k)}$$

- m_2 – Inverse Generational Distance (IGD): The average distance (over R optimization runs) between the team's predicted front and PFbsf:

$$m_2^{(k)} = \frac{1}{R} \sum_{j=1}^{R} \text{IGD}_j^{(k)}$$

- m_3 – Hypervolume (HV): The average (over R runs) of the dominated volume in objective space, using as reference point the extrema of PFbsf:

$$m_3^{(k)} = \frac{1}{R} \sum_{j=1}^{R} \text{HV}_j^{(k)}$$

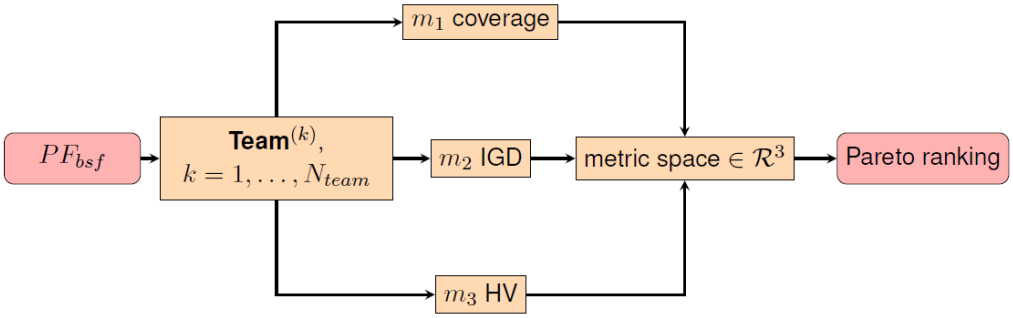

Fig. 2. Flowchart of the metric computation and ranking definition

Each team was represented as a point in the metric space (m_1, m_2, m_3), and the following procedure was used to assign the final ranking:

1) Pareto ranking: A Pareto ranking was performed in the metric space, with the direction of optimization set as: maximize m_1, minimize m_2, and maximize m_3.
2) Front-wise intra-metric ranking: Within each front, teams were ranked individually for each metric using: descending order for m_1 and m_3, ascending order for m_2.
3) Average positional score: For each team k, a scalar score $s^{(k)}$ was computed as:

$$s^{(k)} = \frac{1}{3} \sum_{i=1}^{3} P_i^{(k)}$$

where $P_i^{(k)}$ is the position of team k in metric i, within its Pareto front.
4) Final ranking assignment: Teams were ordered first by front number (lower is better), then by ascending $s^{(k)}$ within the same front. Ties in $s^{(k)}$ were assigned the same rank.

The flowchart of the metric computation and ranking definition is shown in Figure 2.

This method ensured a fair and balanced evaluation, rewarding models with consistent performance across metrics while maintaining diversity among modeling strategies. The same ranking procedure was also applied to the Novelty track, where the metric space was defined by the qualitative scores assigned by the Advisory Board for Novelty, Design insight, and Efficiency.

IV. Procedures

This section describes the procedures presented by the teams for predicting motor performance in the Interpolation and Extrapolation tasks. Table III summarizes the approaches and the key features for each.

Altair E-Motor Team (France), Team 1

The Altair E-Motor Team adopted a neural-network-based reduced-order modeling workflow [18] implemented in *Altair Twin Activate* using the RomAI tool. The software imported the CSV datasets (12 inputs, 7 outputs) and trained a single multi-output network for Motor A with an 80/20 train–validation split (4096 samples). Model accuracy was evaluated by loss reduction and a correlation-based quality index close to 1. A unified dataset (8448 samples from Motors A, B, C) was also created, with 200 A, 200 B, and 50 C samples reserved for testing. Since the combined "ABC" model slightly reduced accuracy on Motor A, two networks were adopted: one trained on Motor A only and one on the full dataset, automatically selected according to the stator outer diameter. The final surrogate was exported as an FMU (Functional Mock-up Unit) for seamless integration within the NSGA-III optimization workflow.

CAD Lab Team (Italy), Team 3

The CAD Lab Team developed seven independent deep neural networks (DNNs) [19], each dedicated to one of the seven output quantities, using 4096 samples from Motor A for the interpolation task. Outliers were manually removed, and model architectures were optimized via *Keras Tuner* [20], exploring up to six hidden layers, 256 neurons per layer, and 20% dropout. The best configuration achieved an overall R^2 of 0.96 and a mean absolute percentage error (MAPE) of 1.28%. For the extrapolation on Motor C, where only 256 samples were available, a structured data augmentation procedure based on Latin Hypercube Sampling (LHS) [21] was applied by adding $\pm$ 1% noise exclusively to the eight design inputs. The resulting training dataset combined Motors A, B, and augmented C samples (100 real + 500 augmented), while the validation set consisted solely of Motor C data (100 real + 500 augmented). This approach ensured that validation points were unseen during training, providing a strict generalization test. The key innovations lie in the physics-informed data augmentation and the Motor C-exclusive validation strategy, which enabled robust extrapolation performance. The final model achieved an overall R^2 of 0.92 and 5.07% MAPE, confirming its ability to generalize under limited-data conditions.

CREATORs (Austria–Germany), Team 4

The CREATORs team developed surrogate models to predict the seven target quantities of the IPM motor datasets using the full data without outlier removal. For the interpolation task, the Motor A dataset was divided into 80% training and 20% validation subsets and repeated across ten random seeds to assess statistical robustness. Four regression methods were in-

TABLE III
SUMMARY OF THE SURROGATE MODELING PROCEDURES ADOPTED BY THE PARTICIPATING TEAMS.

Team	Approach and Key Features
1 - Altair E-Motor Team (France)	Neural-network reduced-order model (RomAI, *Altair Twin Activate*); two multi-output NNs (Motor A / ABC) automatically selected by stator outer diameter; FMU export for NSGA-III integration.
3 - CAD Lab Team (Italy)	Seven DNNs (one per output) optimized via *Keras Tuner*; Latin Hypercube Sampling (±1 %) for data augmentation on scarce Motor C dataset; C-exclusive validation ensuring strict generalization; achieved 0.92 R^2 and 5 % MAPE.
4 - CREATORs (Austria–Germany)	Gaussian Process and Polynomial Chaos Kriging surrogates; Motor A trained with 80/20 splits over 10 random seeds; Motor C modeled by retraining GP/PCK on 200 samples; consistent accuracy across electromagnetic and thermal targets.
7 - ECAD (South Korea)	Multilayer Perceptron surrogate with Bayesian hyperparameter optimization; Deep Transfer Learning from combined Motors A–B to fine-tune on 256 Motor C samples; improved accuracy except for Von Mises stress.
8 - Electa (Belgium)	Neural-network surrogate in *PyTorch* tuned via *Optuna*; ensemble of general and task-specific NNs for torque ripple and stress; transfer learning from Motors A–B to Motor C with full fine-tuning and adaptive learning rate.
12 - GTB ULille (France)	Ensemble of fully connected NNs with input selection by correlation analysis; constraint-aware data resampling; transfer learning to Motor C by replacing and fine-tuning final layers with staged learning rates for improved generalization.
13 - Sonceboz Group (Switzerland)	Conditional Variational Autoencoder (CVAE) surrogate using *AIXD*/*PyTorch*; 9D latent space regularized via reparameterization; six-block MLP encoder–decoder enabling forward and inverse motor design.
14 - IEML (Iran)	Sequential Bayesian Regularized TLBO (SBRTO) combining MLPANN and TLBO; NSGA-II feature selection; cross-validated initialization for faster convergence; physics-inspired dataset conversion for extrapolation.
16 - LASEteam (Canada)	Hybrid predictor–corrector framework combining GPR, XGBoost, and MLP; diversity subsampling for balanced training; corrector trained on residuals to minimize extrapolation error and improve robustness.
18 - ManTriS (Italy)	Hierarchical GPR–decision tree surrogate; anisotropic Matérn 5/2 kernel with MLE tuning; seven single-output GPRs per motor type; 1256-sample training (A,B,C); achieved $R^2 > 0.95$ for most KPIs.
19 - McGill MagNets (Canada)	Hybrid framework using CVAE-generated synthetic data and FNN; CVAE trained with reconstruction and KL loss to preserve input–output correlations; FNN pretrained on synthetic data then fine-tuned on real samples for improved generalization.
20 - MELSUR (Japan–USA)	Physics-assisted ANN framework with seven two-layer models (one per output); two-stage training using Motors A–B and fine-tuning on C; automatic hyperparameter and weight-decay optimization via *Optuna*.
21 - MLotors (India)	Set of seven dense NNs (GELU activation, AdamW) trained per output on Motor A; hybrid extrapolation with transfer learning and residual correction; separate physics-informed NN for torque ripple using 19 engineered features.
25 - SHIME-PARFAIT (Japan)	Deep Operator Network (DeepONet) surrogate with five-fold ensemble and *Optuna* tuning; sequential transfer learning (A→B→C) updating only output layers; achieved $R^2 > 0.99$ and strong cross-domain generalization.
26 - The Overfitters (Austria)	Hybrid XGBoost–FNN surrogate; correlation-based feature selection with logarithmic scaling; XGBoost features fed to 3-layer FNN (128–64–32 ReLU); achieved R^2 of 0.98–0.99 (temperature) and 0.90–0.94 (stress).

vestigated: Polynomial Chaos Expansion (PCE) [22], Gaussian Process (GP) [23], Polynomial Chaos Kriging (PCK) [24], and Feedforward Neural Networks (FNN) [25]. Model accuracy was evaluated using mean and maximum Absolute Percentage Error (APE) across all outputs. GP models achieved the best accuracy for torque and torque ripple, while PCK outperformed for stress- and temperature-related quantities. For extrapolation on Motor C, three strategies were examined: a linear combination of pre-trained interpolation models from Motors A and B, calibrated with 200 Motor C samples; new GP and PCK models trained exclusively on the 200 Motor C samples, which yielded the best performance; and a FNN trained on combined A–B–C data including augmented C samples. Fifty-six Motor C samples were kept for independent evaluation. Results showed high interpolation accuracy for most outputs and lower, but still consistent, performance in extrapolation due to limited data, with pure Motor C models unexpectedly outperforming transfer-learning schemes.

ECAD (South Korea), Team 7

The ECAD team implemented a multilayer perceptron (MLP) surrogate model whose hyperparameters were optimized through Bayesian search. For interpolation, Datasets A and B (4096 samples each) were jointly used for training and validation, achieving mean errors around 1% for most output quantities. For extrapolation on Motor C, where only 256 samples were available, the team adopted a Deep Transfer Learning (DTL) strategy [26], [27]. The base network pretrained on the combined A–B data provided initial weights that were fine-tuned using the small Motor C dataset, allowing effective adaptation of the learned representations to the new motor geometry. This approach substantially improved prediction accuracy for Motor C across most outputs, except for Von Mises stress, and demonstrated the suitability of DTL-based surrogate modeling for scalable optimization of various traction motor topologies.

Electa (Belgium), Team 8

The Electa team developed a neural-network surrogate in *PyTorch* to predict motor performance metrics from geometric and supply parameters. For interpolation, the Motor A dataset was standardized and split into 80% training and 20% validation subsets. Hyperparameters were optimized via *Optuna*, leading to an architecture with 8 inputs, two hidden layers of 72 neurons using Tanh activation, and 7 outputs. Training used the Adam optimizer (learning rate 0.00014, MSE loss, batch size 16) for up to 10 000 epochs with early stopping after 100 epochs of stagnation. Because torque ripple and Von Mises stress were harder to predict, two specialized networks were trained for these quantities and combined with the general model into an ensemble. For extrapolation, transfer learning was employed. A base model was pretrained on merged Motor A + B datasets (12 inputs, 5 hidden layers with 112 neurons, Sigmoid activation) and fine-tuned on the limited Motor C data without freezing layers, enabling full adaptation. Fine-tuning used a lower learning rate (0.00036), batch size 32, and early stopping (patience 50). All parameters were selected automatically using *Optuna*. This full retraining approach ensured flexibility and robustness in handling unseen configurations.

GTB ULille (France), Team 12

The GTB ULille team developed a surrogate modeling framework based on ensembles of fully connected neural networks (NNs) to map motor design parameters to performance indicators. For interpolation, two NN ensembles were trained separately for Motors A and B. Each ensemble contained seven individual networks, each predicting one output quantity, with input features selected through correlation analysis to reduce redundancy. A resampling procedure was applied to the training data, duplicating samples that satisfied design constraints to improve model balance and stability.

For extrapolation, a transfer learning strategy was implemented to adapt the NN ensemble to the limited Motor C dataset. The pre-trained networks from Motors A and B were used as base models. The last two layers of each NN were replaced with layers containing fewer neurons to enhance generalization. Fine-tuning proceeded in two stages: first, only the new layers were trained using a high learning rate while freezing the others; then, all layers were unfrozen and retrained with a lower learning rate. The final surrogate function was obtained by aggregating the ensemble predictions, ensuring improved adaptability under low-data conditions.

Sonceboz Group (Switzerland), Team 13

The Sonceboz Group proposed a deep-learning surrogate based on a Conditional Variational Autoencoder (CVAE) architecture. The encoder maps eight geometric design parameters to seven output quantities, forming a latent representation that serves as the forward surrogate model. The decoder reconstructs or generates new motor geometries from target performance quantities, thus enabling inverse design. The latent space has nine dimensions and is regularized toward a multivariate Gaussian distribution through the reparameterization trick. The implementation, described in [28], was developed in Python using the *AIXD* library [29], built on *PyTorch*. Both encoder and decoder consist of six multilayer perceptron (MLP) blocks with leaky-ReLU activations, batch normalization, and dropout (0.1) to prevent overfitting. This architecture provides a compact bidirectional representation of motor geometry–performance relationships.

IEML (Iran), Team 14

The IEML team developed a sequential machine-learning framework, named Sequential Bayesian Regularized Teaching– Learning- Based Optimization (SBRTO), which integrates a Bayesian regularization-enhanced multilayer perceptron artificial neural network (MLPANN) with the Teaching–Learning-Based Optimization (TLBO) algorithm [30]. In the first stage, cross-validated MLPANNs are used to initialize the TLBO population, reducing the computational cost associated with random initialization and accelerating convergence toward the global optimum. Cross-validation ensures that the models learn the underlying data structure rather than memorizing it, thus preventing overfitting. TLBO was selected for its parameter-free nature, which improves training efficiency and eliminates the need for additional tuning variables.

A multi-objective feature selection step based on NSGA-II was introduced before training to identify the most relevant design parameters across physical domains, aligning the surrogate construction with the multi-physics nature of the motor problem. For the extrapolation task, an innovative dataset-conversion procedure was implemented, acting as a computational-intelligence counterpart to traditional motor sizing equations. The resulting hybrid SBRTO model demonstrated consistent predictive performance and improved interpretability through the integration of domain-informed learning mechanisms [31].

LASEteam (Canada), Team 16

The LASEteam developed a hybrid predictor–corrector modeling framework combining multiple machine learning algorithms to enhance interpolation and extrapolation accuracy. Training, validation, and test sets were generated using the diversity subsampling algorithm [32], ensuring uniform coverage of the input domain. The predictor model was constructed as a synthesis of three complementary algorithms: Gaussian process regression (GPR), extreme gradient boosting (XGB), and multilayer perceptron (MLP), each selected for its individual strength in modeling specific outputs. A separate corrector stage was then trained using the same composite architecture on the residuals of the predictor, estimating extrapolation errors and providing corrective adjustments. The combined predictor–corrector structure reduced bias and improved robustness in low-data extrapolation scenarios.

ManTriS (Italy), Team 18

The ManTriS team implemented a hierarchical surrogate modeling strategy combining Gaussian Process Regression (GPR) with a decision-tree framework to predict the seven key performance indicators (KPIs) of the three IPM motor families. The GPR models [33] were based on an anisotropic Matérn 5/2 kernel, offering flexibility across heterogeneous design parameters. Kernel variance, length scales, and noise variance were tuned through maximum-likelihood estimation (MLE) following [34], minimizing the negative log-likelihood to identify optimal hyperparameters. Because GPR is inherently

single-output, a separate model was trained for each KPI. The decision-tree structure automatically partitioned the design space, selecting the most appropriate GPR configuration for each motor type (A, B, or C) according to input ranges and geometric characteristics. This hierarchical approach improved robustness and prevented overfitting while maintaining computational efficiency. Training was performed in MATLAB using the `fitrgp` function with 1256 samples (500 from Motor A, 500 from Motor B, and all 256 from Motor C). Twenty-one GPR models were trained in about 60 seconds. The surrogate achieved R^2 scores above 0.95 for most outputs, with the worst and the best accuracy obtained for torque ripple ($R^2 = 0.82$–0.84) and copper mass ($R^2 = 1.00$), respectively, confirming strong interpolation and extrapolation consistency across all physical domains [35], [36].

McGill MagNets (Canada), Team 19

The McGill MagNets team proposed a hybrid data-driven framework that combines real and synthetic data to improve surrogate model generalization. A Conditional Variational Autoencoder (CVAE) was used to generate realistic synthetic samples while preserving the statistical relationships between input and output features [37], [38]. The CVAE architecture includes an encoder that maps paired input–output samples $[x, y]$ into a latent representation defined by a mean and variance, and a decoder that reconstructs outputs from latent variables and the original inputs. The network was trained using a combination of reconstruction loss and Kullback–Leibler divergence to ensure meaningful latent representations. Once trained, the CVAE could generate new synthetic outputs by sampling from a standard normal latent distribution and decoding with any input x. For improved extrapolation, additional samples were drawn from rarely populated latent regions or slightly perturbed inputs. A Feedforward Neural Network (FNN) was first pretrained on the CVAE-generated data to capture global patterns and then fine-tuned on the real dataset to adapt to measured distributions. This two-stage training approach reduced overfitting, improved robustness, and accelerated convergence, while synthetic data were shown to match the true data distribution closely, confirming the CVAE's physical consistency.

MELSUR (Japan–USA), Team 20

The MELSUR team implemented a physics-assisted artificial neural network (ANN) framework composed of seven independent two-layer models, each dedicated to predicting one output quantity. This structure allowed each network to specialize in its corresponding physical variable. Domain knowledge of motor physics was incorporated by identifying direct or proportional relationships between certain input and output parameters and including these derived features as additional inputs.

A two-stage training strategy was adopted to enhance generalization on Motor C. During pretraining, all available data from Motors A, B, and 70% of C were used to build a generalized model. In the second stage, fine-tuning was performed exclusively on Motor C data to refine the predictions while retaining the learned global trends. Hyperparameters—such as hidden-layer size, learning rate, number of epochs, and weight-decay coefficient—were automatically optimized via the *Optuna* library [39]. Weight decay was applied as a regularization term to prevent overfitting and promote smoother convergence. This automated and physics-informed procedure ensured stable performance across both interpolation and extrapolation tasks.

MLotors (India), Team 21

The MLotors team developed separate neural models to predict each of the seven output quantities: torque, torque ripple, copper and magnet mass, power factor, Von Mises stress, and temperature. Each model comprised six densely connected layers with Gaussian Error Linear Unit (GELU) activations and decreasing neuron counts from 512 to 32. Models were trained and tested on an 80/20 split of normalized Motor A data using the AdamW optimizer and mean-squared-error loss. This configuration achieved mean absolute errors below 1% for most metrics, except for Von Mises stress (2.7%) and torque ripple (6.2%).

For extrapolation, a hybrid scheme combining neural networks, transfer learning, and residual correction was implemented. A bi-layer dense network with 64 neurons per layer and a linear output head was pretrained on Motors A and B, then fine-tuned on Motor C data. Residual learning, inspired by XGBoost, was used to refine predictions. For torque ripple, a separate physics-informed neural network (PINN) was introduced, using 19 derived input features to capture domain-specific correlations and enhance generalization to unseen motor configurations.

SHIME-PARFAIT (Japan), Team 25

The SHIME-PARFAIT team developed a surrogate modeling framework based on Deep Operator Networks (DeepONets) [40], [41], combined with ensemble learning and a sequential transfer-learning procedure. For interpolation, a multi-output DeepONet was trained using seven selected input features to predict seven physical outputs, including torque, thermal quantities, and mechanical stress. Robustness and accuracy were enhanced through five-fold ensemble training [42], where each submodel was independently tuned via *Optuna*. The resulting ensemble achieved R^2 values exceeding 0.99 across all outputs, demonstrating strong stability and consistency. For extrapolation on the limited Motor C dataset (256 samples), a two-stage transfer learning strategy [43] was adopted. The base model was first pretrained on Motor A (4096 samples), fine-tuned on Motor B (4096 samples), and finally adapted to Motor C. During each transfer, only the output layers were updated while hidden layers remained frozen, allowing efficient reuse of learned physical representations. This sequential adaptation reduced prediction error and variance, yielding an accurate and computationally efficient surrogate applicable to multi-physics motor design.

The Overfitters (Austria), Team 26

The Overfitters team proposed a hybrid surrogate framework combining Extreme Gradient Boosting (XGBoost) and Feedforward Neural Networks (FNNs) to predict key performance indicators—torque, temperature, and Von Mises stress—from geometric parameters. Input features were first selected by correlation analysis and logarithmically transformed to improve linearity. XGBoost captured structured dependencies and produced intermediate features that were fed into a se-

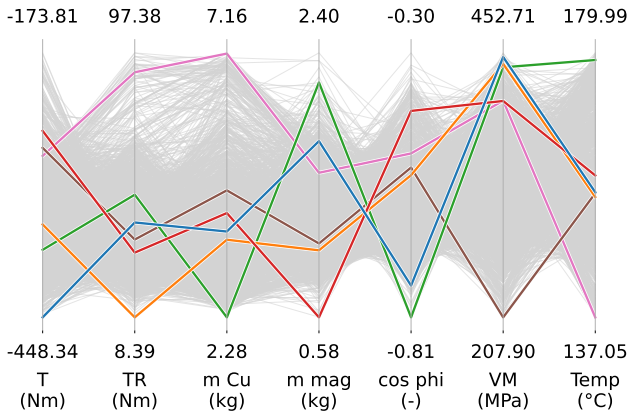

Fig. 3. Motor A. Parallel Coordinate Plot of the best Pareto front so far. Over the light gray envelope showing all configurations, the configurations corresponding to the best value in each objective are highlighted.

quential FNN with three hidden layers (128–64–32 neurons, ReLU activations, dropout). The model was trained using the Adam optimizer with exponential learning-rate decay and early stopping to avoid overfitting. For torque prediction, original features were concatenated with XGBoost outputs to capture higher-order nonlinearities. The resulting hybrid model achieved strong generalization, with R^2 between 0.98–0.99 for temperature, 0.90–0.94 for stress, and 0.67–0.70 for torque.

V. Contest Results

The variety of adopted approaches highlights the methodological richness of the GalFer Contest. The following section reports the quantitative results obtained under uniform evaluation conditions.

A. Prediction on Motor A

The main result on the motor performance is the Pareto Front best so far, which contains all the values reached by the optimizer in its quest for the multi-objective optimum. As an example of the results reached by the procedure, a Parallel Coordinate Plot (PCP) of the best so far Pareto Front obtained for motor A is shown in Figure 3.

These values are validated by the FEM procedure. On the whole set of PFbsf shown as a light gray background, the values for each of the configuration reaching the minimum value in each of the objectives are highligthed. This plot gives an indication of the trade-off among each of the objectives.

For what concerns accuracy of the surrogate models, Figure 4 compares the predicted and FEM-validated values, belonging to each team PF, obtained by all teams for the seven target quantities in the Interpolation task. The scatter distributions confirm that most surrogate models successfully generalized to these unseen regions of the design space, since these configurations were not included in the training datasets, as they correspond to Pareto-optimal design region that is sparsely represented, or even unseen, within the original sampling. The prediction spread was smallest for torque and power factor, while larger deviations appeared for torque ripple and Von Mises stress, which are intrinsically more sensitive to geometric details and local nonlinearities. Temperature and material-mass predictions also showed systematic consistency, confirming that data-driven surrogate models can approximate coupled multi-physics behaviors beyond their nominal training domain when sufficient data coverage is ensured.

Figure 5 reports the relative prediction error for each team, computed on the points of their respective PFs after FEM validation. For every output quantity, the columns represent the 95th-percentile error, while the triangular markers indicate the average value. The overall distribution confirms the consistency of the surrogate predictions across different modeling strategies: in most cases, both mean and high-percentile errors remain limited, showing that surrogate models can reliably approximate the true multi-physics response even in the Pareto-optimal region, which was not explicitly used for training.

B. Prediction on Motor C

Also in this case the PFbsf is reported as a PCP and it is shown in Figure 6, analogous to the one already described for motor A. Figures 7 and 8 illustrate the prediction performance of the participating teams in the Extrapolation task. As in the Interpolation case, Figure (7) compares the predicted and FEM-validated quantities on the teams Pareto Fronts, while Figure 8 reports the relative errors, with columns indicating the 95th-percentile and triangles the mean values for each output.

The plots reveal that, despite the extremely limited number of available samples for Motor C, most models retained good predictive capability. The correlation between predicted and true values remains strong across the majority of performance metrics, confirming that the learned mappings from Motors A and B generalized effectively to a motor configuration poorly represented in the training data.

The error analysis in Figure 8 further highlights that, for many teams, both average and 95th-percentile errors stay within moderate bounds even under this low-data regime. Overall, the results demonstrate that the proposed benchmark successfully discriminates between interpolation accuracy and true generalization capability, providing a realistic test for data-driven design methods under data scarcity.

C. Geometric variation

The results presented in the PCP graphs contain the variation of the configurations in the objective space. It is nevertheless interesting to show how these configurations are changing in the design variable space. As it is too complex to show it for many motor structures, here three geometries corresponding to key performance are shown:

- maximum torque;
- minimum temperature;
- minimum Von Mises.

The geometries are presented in Figure 9 and they show how the optimizer was able to find magnetic and mechanical structures able to fulfill the maximum performance in each objective. Obviously, these are extreme points of the PFbsf, and none of them is a real compromise in different design requirements; anyway, they are examples of the physical validity of the optimization process.

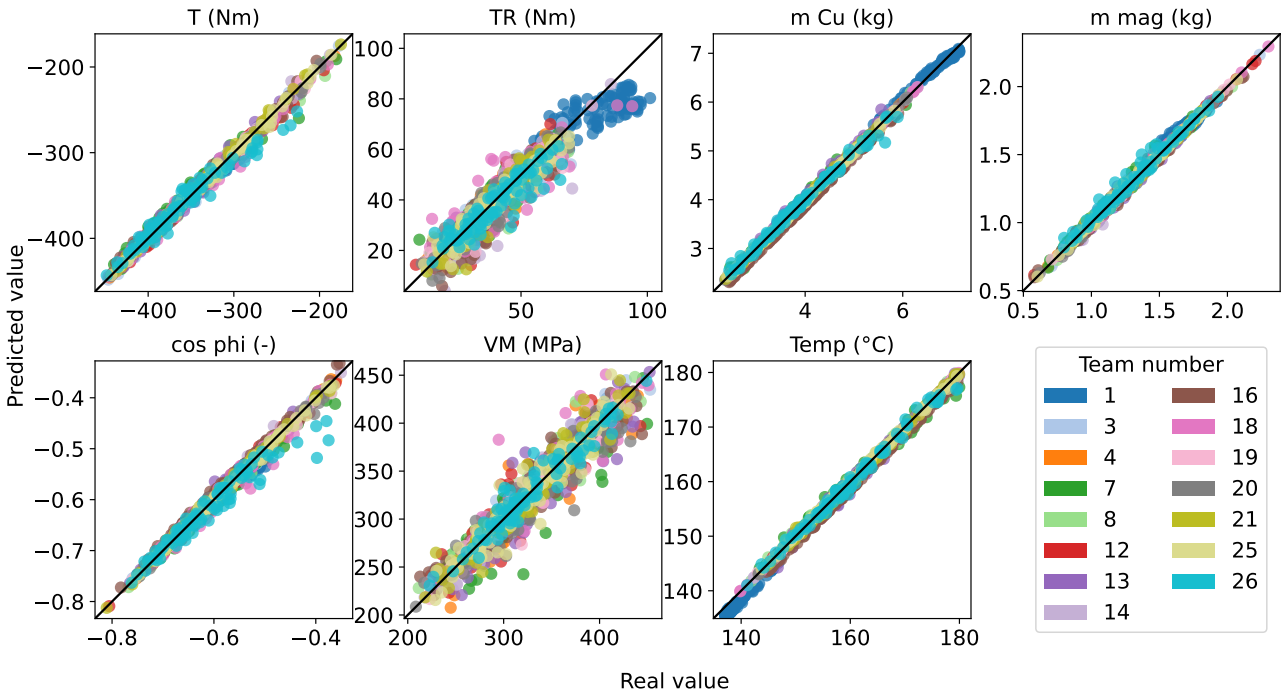

Fig. 4. Interpolation track: Real vs. Predicted chart for each of the objectives computed.

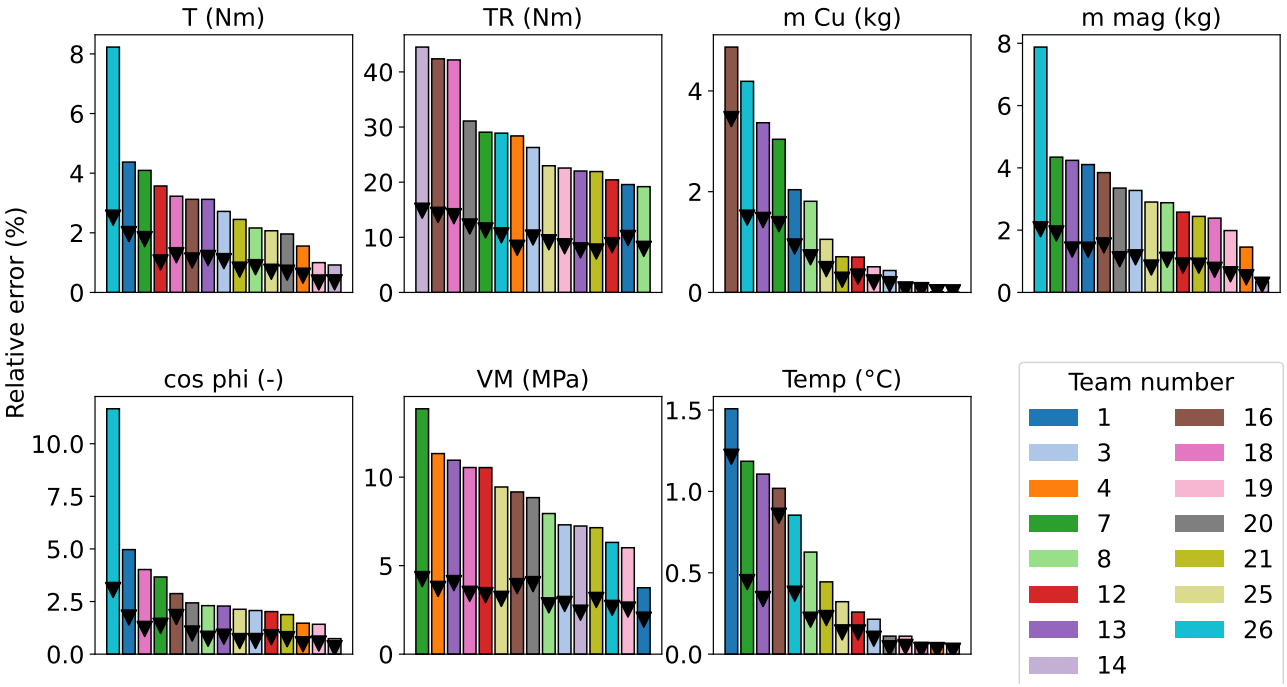

Fig. 5. Interpolation track: 95-th percentile error between surrogate model results and true FEM evaluated values, for each team ordered by frequency. The symbol represents the average error.

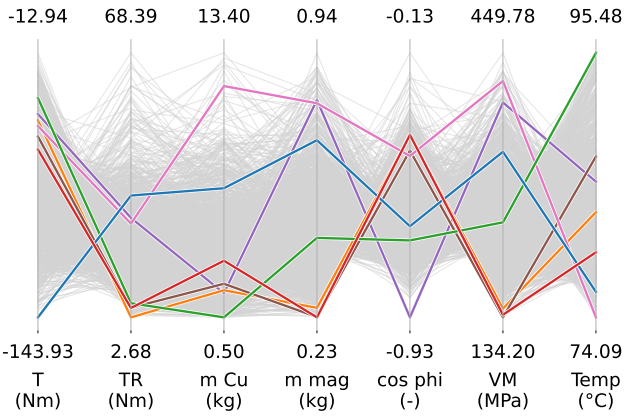

Fig. 6. Motor C. Parallel Coordinate Plot of the best Pareto front so far. Over the light-gray envelope showing all configurations, the configurations corresponding to the best value in each objective are highlighted.

VI. Discussion

A. Outline of results and ranking

The GalFer Contest established a unified framework for assessing data-driven modeling in the design of traction electric motors, linking open multi-physics data generation with systematic benchmarking of surrogate models. The open dataset, combining electromagnetic, thermal, and structural simulations, for three IPM motor families, proved to be a robust and coherent foundation for surrogate model training and validation, enabling reproducible evaluation of model accuracy, generalization, and innovation.

An analysis of the procedures submitted by the teams show how a large majority of them, 9 over 15, are based on the neural network paradigm, often hybridizing the deep learning scheme with specific pre-processing and augmentation of datasets. Gaussian Process was used by 3 teams and again here the basic technique was enriched with peculiar features. Two teams used Conditional Variational Autoencoder methodology. Only one team was explicitly using of physics-assisted neural networks.

The metric values obtained by the ranking procedure outlined in Section III are presented in Table IV, both on motor A and C.

Based on these values, the final ranking of teams could be performed. All teams were ranked according to the Pareto dominance criterion defined in Section III, and grouped into non-dominated sets. Submissions in the first non-dominated set were assigned PF rank 0, those dominated only by front-0 submissions were assigned PF rank 1, and so forth. Lower PF ranks, therefore, indicate better global trade-offs among the three metrics (with 0 denoting the best PF). The resulting PF index, ranging from 0 to 6, is reported in Table V.

The *Interpolation* track was won by the industrial team *Melsur (20)* that used physics-assisted neural networks and by academic team *CREATORS (4)* using Gaussian process methodology. The *Extrapolation* track saw in the first place the academic team *ManTris (18)* adopting again the Gaussian process regressor numerical scheme.

TABLE IV
METRIC VALUES FOR THE ALL TEAMS BOTH ON MOTOR A AND C. THE PERFORMANCE OF THE ORIGINAL DATASETS ARE LISTED A $team0$.

Team	Motor A			Motor C		
	Cov.	HV	IGD	Cov.	HV	IGD
0	**85**	**0.2251**	**0.1250**	**74**	**0.1760**	**0.2177**
1	99	0.6494	0.0150	52	0.1810	0.2070
3	93	0.1992	0.1772	96	0.1601	0.2427
4	96	0.1974	0.1728	97	0.1521	0.2568
7	96	0.1997	0.1692	100	0.1571	0.2577
8	92	0.2051	0.1720	96	0.1511	0.2553
12	92	0.2024	0.1697	96	0.1493	0.2411
13	94	0.1971	0.1652	95	0.1757	0.2314
14	95	0.1936	0.1728	94	0.1683	0.2501
15	93	0.1959	0.1732	89	0.1544	0.2521
18	91	0.1860	0.1683	96	0.1538	0.2612
19	92	0.1907	0.1698	95	0.1544	0.2535
20	93	0.1888	0.1750	93	0.1558	0.2688
21	96	0.2056	0.1729	92	0.1518	0.2568
25	92	0.1965	0.1716	95	0.1554	0.2433
26	90	0.1964	0.1637	89	0.1638	0.2314

TABLE V
FINAL RANKING OF THE TEAMS BASED ON THE METRIC VALUES OBTAINED IN BOTH INTERPOLATION AND EXTRAPOLATION TRACKES. PF IS THE PARETO FRONT NUMBER IN THE RANKING.

Rank	Motor A			Motor C		
	Team	PF	Model	Team	PF	Model
1	**20**	**0**	PINN	**18**	**0**	GPR
2	**4**	**0**	GPR, PCK	**4**	**0**	GPR, PCK
3	**21**	**0**	DNN, PINN	**8**	**0**	DNN+TL
4	3	0	DNN	12	0	DNN+TL
5	14	0	DNN	7	0	DNN+TL
6	18	0	GPR	20	0	PINN
7	1	0	ROM-NN	21	0	DNN, PINN + TL
8	19	1	CVAE+DNN	19	1	CVAE+DNN
9	7	1	DNN	16	1	GPR, DNN
10	13	1	CVAE	3	1	DNN
11	8	2	DNN	25	2	deepONet+TL
12	25	2	deepONet	14	2	DNN
13	26	2	DNN	13	3	CVAE
14	12	3	DNN	26	4	DNN
15	16	4	GPR, DNN	**0**	**5**	–
16	**0**	5	–	1	6	ROM-NN

Anyway, all the surrogate models were able to produce reliable results when linked in the optimization loop. It must be remarked that, notwithstanding that the datasets were generated trying to have a good uniform sampling distribution on the whole design variable space, the surrogate models were able to give well founded estimates on the Pareto Front that is a partial subset of the space.

B. Innovation

A key outcome of the GalFer Contest is the assessment of *innovation*, understood as the capability of surrogate-based optimization to discover Pareto-optimal configurations beyond the resolution of the original dataset. While the dataset itself provides a uniform sampling of the design space, its points can only occasionally belong to the region of the PFbsf. In contrast, when the surrogate models were embedded in the multi-objective optimization pipeline and their results were validated through FEM analysis, they consistently identified the PFbsf region.

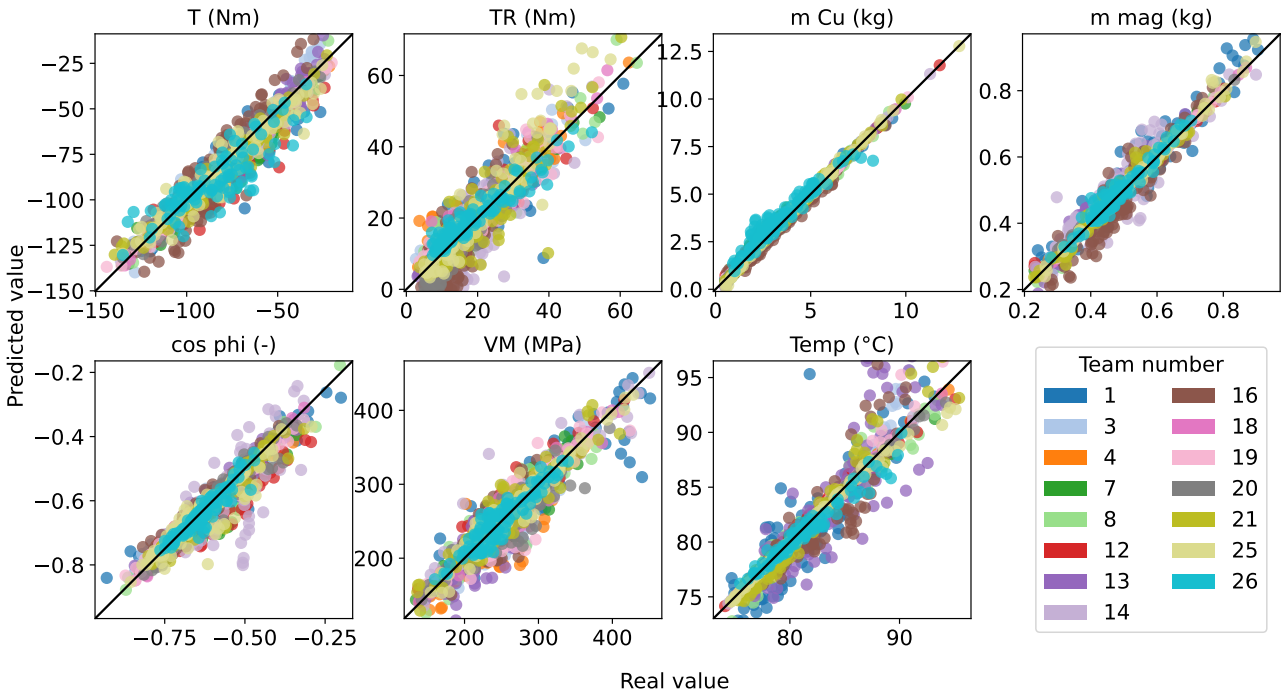

Fig. 7. Extrapolation track: Real vs. Predicted chart for each of the objectives computed.

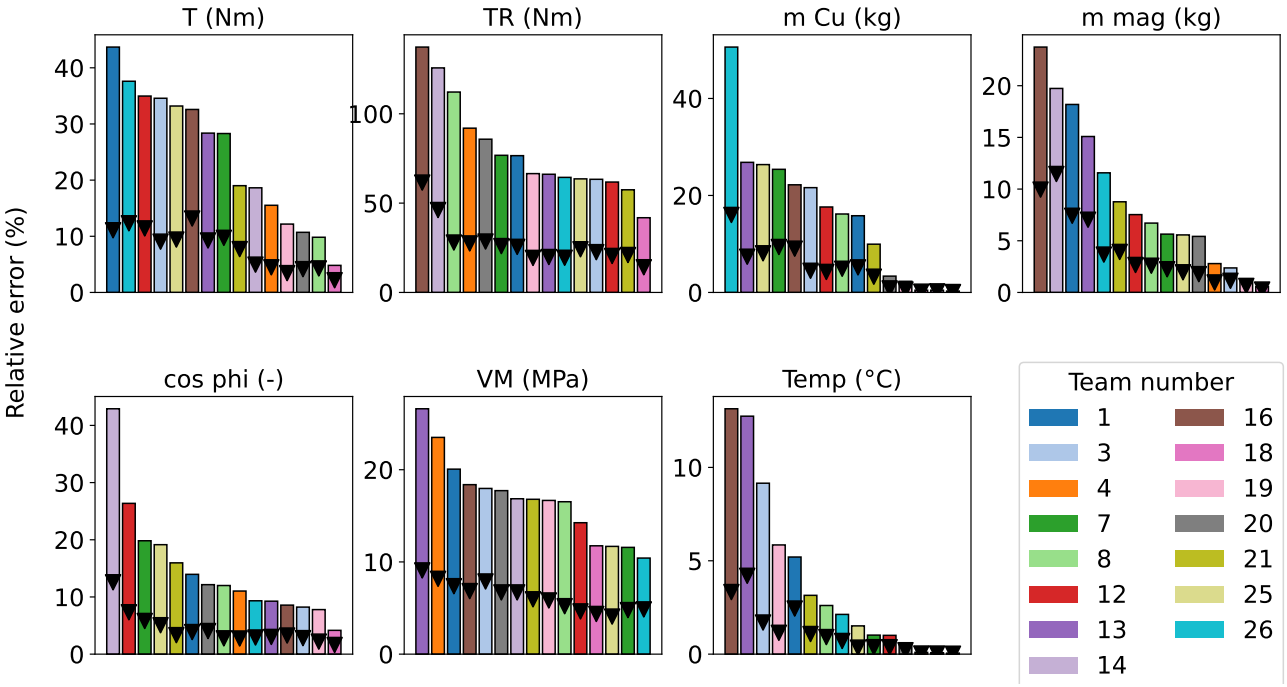

Fig. 8. Extrapolation track: 95-th percentile error between surrogate model results and true FEM evaluated values, for each team ordered by frequency. The symbol represents the average error.

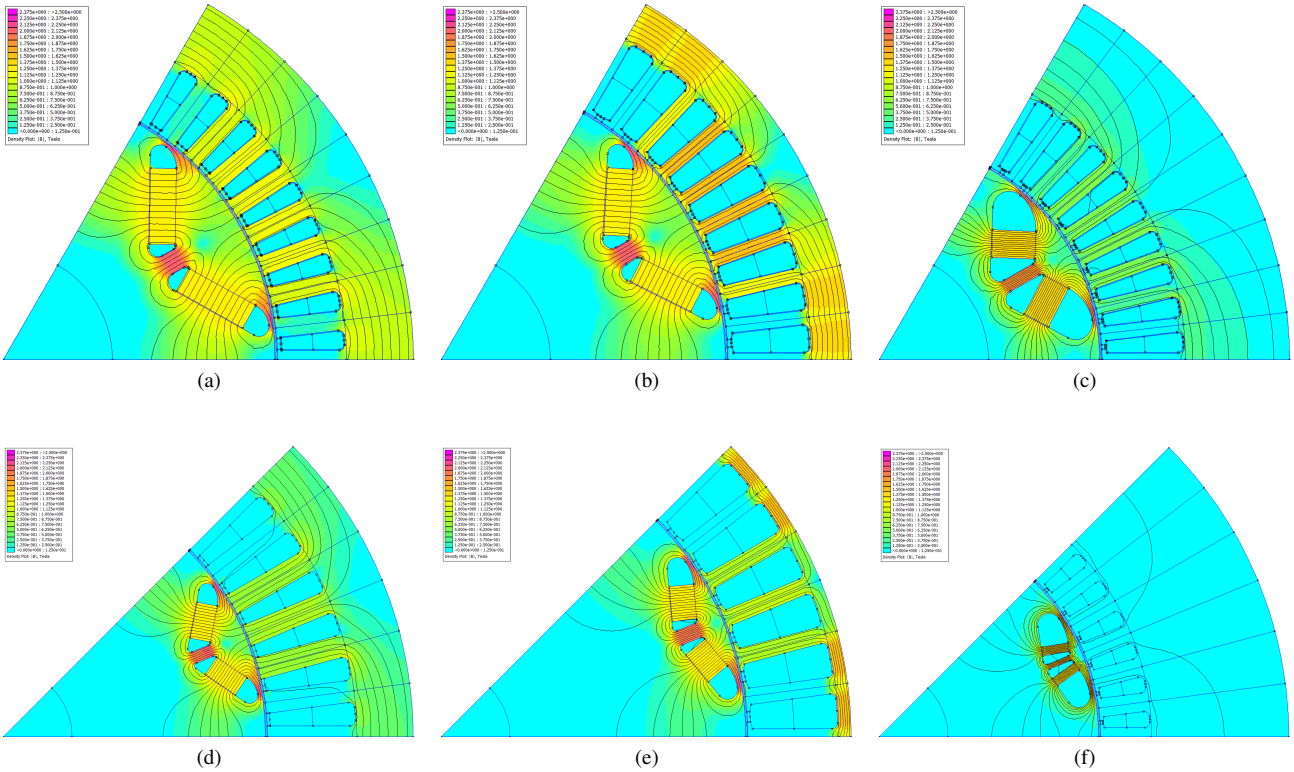

Fig. 9. Cross sections of motor A (upper row) and C (lower row) configurations corresponding to three peculiar features. From left to right: highest torque, lowest temperature, and lowest Von Mises stress values.

This improvement is quantitatively confirmed by the metrics and rankings presented in Table IV) and V, respectively, where the dataset is described as Team 0. In both the Interpolation and Extrapolation tracks, the original dataset itself is systematically dominated by surrogate results, indicating that the direct FEM-sampled points could not reach the optimal trade-offs revealed by the surrogate-driven searches. Conversely, the top-performing models exhibited larger Coverage, lower IGD, and higher HV, highlighting their ability to generalize the physics-informed relationships learned from data into truly optimized motor designs. The comparison of the team results and the dataset in the metrics space is also well illustrated in Figures 10 and 11, where the marker position indicates the HV and the IGD, while the size of the marker is proportional to the Coverage. In both figures, the dataset itself lies systematically in the lowest-performance region of the metric space, while the surrogate-based optimizations occupy zones of higher hypervolume and lower IGD. This displacement clearly visualizes the innovation capability of the surrogate models, defined as their ability to discover Pareto-optimal regions that extend beyond the resolution of the original training data. The consistent separation between the dataset and the best-performing teams demonstrates that the use of learned surrogates within an optimization loop enables exploration of new, physically consistent design trade-offs, even under data scarcity as in the Motor C case.

It is worth noting that, if the same optimization would have been performed directly through FEM analysis, the resulting Pareto front would naturally be accurately reconstructed, but at the cost of an extremely high computational burden. In contrast, the surrogate-based workflow achieves comparable insight into the optimal design region at a negligible fraction of the computational cost, demonstrating its effectiveness as a scalable alternative for multi-physics optimization.

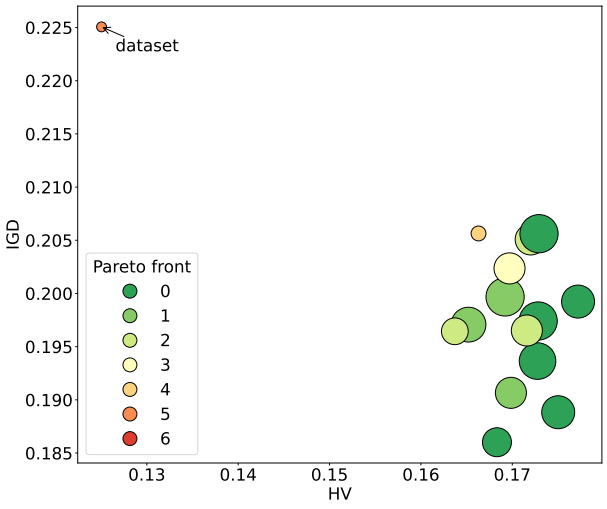

Fig. 10. Motor A: scatter plot of the results of the procedures in the metric space. The position of the marker gives the hypervolume (HV) and the inverse generational distance (IGD), while the size of the marker is proportional to the coverage metric. Colors are assigned on the basis of the Pareto ranking.

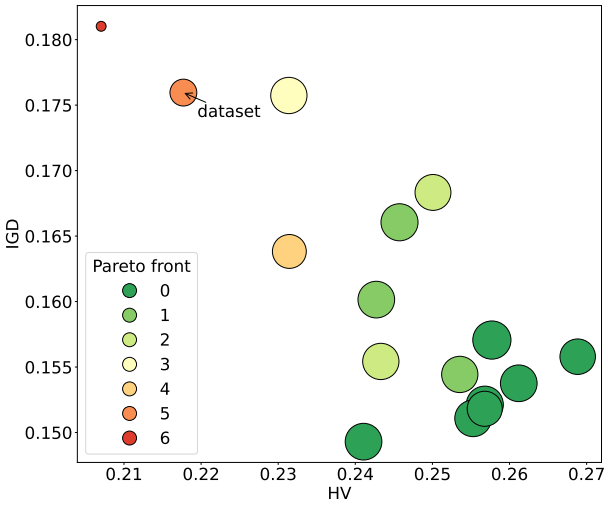

Fig. 11. Motor C: scatter plot of the results of the procedures in the metric space. The position of the marker gives the hypervolume (HV) and the inverse generational distance (IGD), while the size of the marker is proportional to the coverage metric. Colors are assigned on the basis of the Pareto ranking.

VII. Conclusions

Even if the present work represents only a preliminary application of data-driven procedures to an industrial design process, some initial conclusions can already be drawn.

The results demonstrated that surrogate models can accurately reproduce the coupled physical behavior of electric motors, achieving good precision while reducing computational cost by several orders of magnitude. When embedded into multi-objective optimization, these models successfully identified Pareto-optimal configurations outside the resolution of the original dataset, highlighting their capability to generalize physical relationships learned from data.

From a methodological perspective, the GalFer Contest also validated the effectiveness of a common optimization and validation workflow. By applying the same NSGA-III and FEM-based evaluation pipeline to all submissions, differences among teams were attributable to modeling choices, ensuring an objective comparison across diverse approaches, from Gaussian processes and polynomial-chaos expansions to neural networks and transfer-learning architectures.

Overall, the GalFer Contest confirmed the complementarity between the dataset and the surrogate models: the dataset provides a reproducible, physics-consistent reference environment, while surrogate models act as scalable tools for exploring and optimizing complex design spaces. This dual contribution defines a pathway toward standardized, data-driven multi-physics design workflows, to be extended in future benchmark editions and applied to broader classes of electromagnetic devices and coupled-domain systems.

The resulting rankings thus summarize the core contribution of the GalFer Contest: the demonstration that high-fidelity, data-driven surrogates, once coupled with an optimization algorithm, can explore and exploit the design space far beyond the explicit content of the training data. This validates their role not only as predictive tools but also as enablers of design innovation, capable of guiding optimization toward physically consistent and previously unsampled high-performance regions. This outcome constitutes the most distinctive contribution of the GalFer Contest and forms the basis for future extensions of benchmark activities.

Acknowledgments

The authors would like to acknowledge the support of the IEEE Magnetic Society Italy Chapter, the International Compumag Society, MathWorks, the Power Electronics Innovation Center (PEIC) of Politecnico di Torino, and Stellantis for their contribution to the organization and support of the GalFer Contest. The authors also wish to thank the members of the Advisory Board for their valuable evaluation activity and scientific guidance throughout the initiative. Advisory Board members were: Prof. Piergiorgio Alotto University of Padova, Italy, Dr. Costanza Anerdi Politecnico di Torino, Italy, Prof. Manfred Kaltenbacher TU Graz, Austria, Prof. Elena Lomonova Eindhoven University of Technology, the Netherlands, Prof. Dave Lowther Mc Gill University, Montreal, Canada, Prof. Kazuhiro Muramatsu Saga University, Japan, Prof. Shahryar Rahnamayan Brock University, St. Catharines, Canada, Prof. Maurizio Repetto Politecnico di Torino, Italy, Prof. Ruth Sabariego KU Leuven, Belgium, Prof. Oliver Wallscheid University of Siegen, Germany.

This initiative would not have been possible without the active participation and commitment of all the research teams worldwide that contributed their methods and insights to the benchmark.

Affiliations

1) Energy Department “Galileo Ferraris”, Politecnico di Torino, Torino, Italy
2) Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Torino, Italy
3) Department of Industrial Engineering, University of Padova, Padova, Italy
4) Altair Innovation Intelligence, Meylan, France
5) Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy
6) Centrum Wiskunde & Informatica, Amsterdam, Netherlands
7) Technische Universität Darmstadt, Darmstadt, Germany
8) Technische Universität Graz, Graz, Austria
9) Siemens AG, Foundational Technologies, Munich, Germany
10) Department of Mechanical, Robotics, and Energy Engineering, Dongguk University, Seoul, Republic of Korea
11) Department of Automotive Engineering (Automotive-Computer Convergence), Hanyang University, Seoul, Republic of Korea
12) Department of Electrical Engineering, Soonchunhyang University, Asan, Republic of Korea
13) Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium - EnergyVille, 3600 Genk, Belgium
14) University of Lille, EEA Department, L2EP Laboratory, France
15) Sonceboz Group, Lausanne, Switzerland
16) Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
17) Laboratory for Structural Efficiency, University of British Columbia, Vancouver, Canada
18) Department of Electronics and Telecommunication, Politecnico di Torino, Torino, Italy
19) Department of Electrical and Computer Engineering, McGill University, Montreal, Canada

20) Department of Computer Science, Iowa State University, Ames, USA
21) Mitsubishi Electric Research Laboratories, Cambridge, USA
22) Advanced Technology R&D Center, Mitsubishi Electric Corporation, Hyogo, Japan
23) School of Electrical and Computer Sciences, Indian Institute of Technology Bhubaneswar, Odisha, India
24) Aoyama Gakuin University, Sagamihara, Kanagawa, Japan
25) Hosei University, Koganei, Tokyo, Japan
26) Hokkaido University, Sapporo, Hokkaido, Japan
27) Independent researcher, Sunagawa, Hokkaido, Japan
28) Department of Electrics/Electronics & Software, VIRTUAL VEHICLE Research GmbH, Graz, Austria
29) Department of Engineering "Yousef Haj-Ahmad", Brock University, St. Catharines, Canada

REFERENCES

[1] "IEEE dataport a research data platform," https://ieee-dataport.org, [Online; accessed 03-Dec-2025].
[2] "Zenodo," https://zenodo.org/, [Online; accessed 03-Dec-2025].
[3] L. Solimene *et al.*, "Data-driven approaches for electromagnetic analysis of traction electrical motors: A proposal for a benchmark problem," in *2024 International Conference on Electrical Machines (ICEM)*, 2024, pp. 1–7.
[4] V. Parekh, D. Flore, and S. Schöps, "Deep Learning-Based Meta-Modeling for Multi-Objective Technology Optimization of Electrical Machines," *IEEE Access*, vol. 11, pp. 93 420–93 430, 2023.
[5] M. H. Mohammadi, V. Ghorbanian, and D. A. Lowther, "A Data-Driven Approach For Design Knowledge Extraction of Synchronous Reluctance Machines Using Multi-Physical Analysis," in *2018 XIII International Conference on Electrical Machines (ICEM)*, Sep. 2018, pp. 479–485, iSSN: 2381-4802.
[6] "Galileo Ferraris' Contest: comparing data-driven methodologies for the multi-physics simulation of traction electrical machines," [Online; accessed 03-Dec-2025]. [Online]. Available: https://github.com/cadema-PoliTO/GalFer_contest
[7] "IEEE Milestones program," [Online; accessed 03-Dec-2025]. [Online]. Available: https://ieeemilestones.ethw.org/Milestones:Ferraris
[8] "International Compumag Society," https://www.compumag.org/wp/, accessed: 2025-10-23.
[9] S. Ferrari *et al.*, "A multiphysics dataset generation procedure for the data-driven modeling of traction electric motors," *IEEE Access*, vol. 13, pp. 54 534–54 546, 2025.
[10] A. Ibrahim *et al.*, "A novel pareto-optimal ranking method for comparing multi-objective optimization algorithms," 2024. [Online]. Available: https://arxiv.org/abs/2411.17999
[11] M. Chen *et al.*, "Magnet challenge for data-driven power magnetics modeling," *IEEE Open Journal of Power Electronics*, vol. 6, pp. 883–898, 2025.
[12] I. M. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," *Zh. Vychisl. Mat. Mat. Fiz.*, vol. 7, no. 4, pp. 784–802, 1967.
[13] S. Ferrari and G. Pellegrino, "SyR-e Synchronous Reluctance – evolution design," https://github.com/SyR-e/syre_public, 2025, [Online; accessed 03-Dec-2025].
[14] "Finite Element Method Magnetics - Version 4.2," https://www.femm.info/wiki/Documentation/, 2025, [Online; accessed 03-Dec-2025].
[15] T. M. Inc., "Partial differential equation toolbox," Natick, Massachusetts, United States, 2025, [Online; accessed 03-Dec-2025]. [Online]. Available: https://www.mathworks.com/help/stats/index.html
[16] S. Ferrari *et al.*, "Multiphysics Datasets for the Data-Driven Modeling of Traction Electric Motors," Nov. 2025. [Online]. Available: https://zenodo.org/doi/10.5281/zenodo.17648857
[17] Y. Tian *et al.*, "PlatEMO: A MATLAB platform for evolutionary multi-objective optimization," *IEEE Computational Intelligence Magazine*, vol. 12, no. 4, pp. 73–87, 2017.
[18] M. E. Abdollahi, N. Vaks, and B. Bilgin, "A multi-objective optimization framework for the design of a high power-density switched reluctance motor," in *2022 IEEE Transportation Electrification Conference and Expo (ITEC)*, 2022, pp. 67–73.
[19] P. Di Barba *et al.*, "Optimal multi-physics synthesis of a dual-frequency power inductor using deep neural networks and gaussian process regression," *Algorithms*, vol. 18, no. 1, 2025. [Online]. Available: https://www.mdpi.com/1999-4893/18/1/10
[20] M. Z. A. Pon and K. K. Krishna Prakash, "Hyperparameter tuning of deep learning models in keras," *Sparklinglight Transactions on Artificial Intelligence and Quantum Computing*, vol. 1, no. 1, May 2021. [Online]. Available: https://doi.org/10.55011/STAIQC.2021.1104
[21] J. Helton and F. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23–69, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832003000589
[22] D. Shen *et al.*, "Polynomial chaos expansion for parametric problems in engineering systems: A review," *IEEE Systems Journal*, vol. 14, no. 3, pp. 4500–4514, 2020.
[23] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
[24] R. Schobi, B. Sudret, and J. Wiart, "Polynomial-chaos-based kriging," *International Journal for Uncertainty Quantification*, vol. 5, no. 2, 2015.
[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[26] S.-H. Park *et al.*, "Deep transfer learning-based sizing method of permanent magnet synchronous motors considering axial leakage flux," *IEEE Transactions on Magnetics*, vol. 58, no. 9, pp. 1–5, 2022.
[27] S.-H. Park, K.-O. Kim, and M.-S. Lim, "Computationally efficient estimation of pwm-induced iron loss of pmsm using deep transfer learning," *IEEE Transactions on Magnetics*, vol. 59, no. 11, pp. 1–5, 2023.
[28] V. Balmer *et al.*, "Design space exploration and explanation via conditional variational autoencoders in meta-model-based conceptual design of pedestrian bridges," *Automation in Construction*, vol. 163, p. 105411, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092658052400147X
[29] "AIXD: AI-eXtended Design," https://gitlab.renkulab.io/ai-augmented-design/aixd/, 2024.
[30] R. Rao, V. Savsani, and D. Vakharia, "Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems," *Computer-Aided Design*, vol. 43, no. 3, pp. 303–315, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010448510002484
[31] T. Sharifi, A. Eikani, and M. Mirsalim, "Heat transfer study on a stator-permanent magnet electric motor: A hybrid estimation model for real-time temperature monitoring and predictive maintenance," *Case Studies in Thermal Engineering*, vol. 63, p. 105286, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214157X24013170
[32] B. Shang, D. W. Apley, and S. Mehrotra, "Diversity subsampling: Custom subsamples from large data sets," *INFORMS Journal on Data Science*, 2023.
[33] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. Cambridge, MA: The MIT Press, 2006.
[34] C. Lataniotis *et al.*, "UQLab user manual – Kriging (Gaussian process modeling)," Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, Tech. Rep., 2022, report UQLab-V2.0-105.
[35] I. Bilionis and N. Zabaras, "Multi-output local Gaussian process regression: Applications to uncertainty quantification," *Journal of Computational Physics*, vol. 231, no. 17, pp. 5718–5746, 2012.
[36] P. Manfredi and R. Trinchero, "A probabilistic machine learning approach for the uncertainty quantification of electronic circuits based on Gaussian process regression," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 41, no. 8, pp. 2638–2651, Aug. 2021.
[37] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in Neural Information Processing Systems*, C. Cortes *et al.*, Eds., vol. 28. Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf
[38] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *CoRR*, vol. abs/1312.6114, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:216078090
[39] T. Akiba *et al.*, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 2623–2631. [Online]. Available: https://doi.org/10.1145/3292500.3330701
[40] L. Lu *et al.*, "Learning nonlinear operators via deeponet based on the universal approximation theorem of operators," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 218–229, 2021. [Online]. Available: https://doi.org/10.1038/s42256-021-00302-5

[41] H. Sasaki *et al.*, "Prediction of motor characteristic maps via deep operator networks for topology optimization," *IEEE Transactions on Magnetics*, vol. 60, no. 12, pp. 1–5, 2024.

[42] J. Hwang and Y. Iwasaki, "Improving efficiency of autonomous material search via transfer learning from nontarget properties," *Science and Technology of Advanced Materials: Methods*, vol. 3, no. 1, p. 2254202, 2023. [Online]. Available: https://doi.org/10.1080/27660400.2023.2254202

[43] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.