Article

Not peer-reviewed version

Optimization of Manufacturing Processes Using AI-Based Advisory Systems: Casting Application

Sofija Milicic , Amir M. Horr * , Stefanie Elgeti , Manuel Hofbauer , Rodrigo Gómez Vázquez

Posted Date: 8 April 2026

doi: 10.20944/preprints202604.0566.v1

Keywords: data models; machine learning; neural network; manufacturing processes; digitalization; digital advisory system

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0566.v1

Article

Optimization of Manufacturing Processes Using AI-Based Advisory Systems: Casting Application

Sofija Milicic [1], Amir M. Horr [1,*], Stefanie Elgeti [2], Manuel Hofbauer [1] and Rodrigo Gómez Vázquez [1]

[1] LKR Light Metals Technologies Ranshofen, Austrian Institute of Technology, 1210 Vienna, Austria
[2] Institute of Lightweight Design and Structural Biomechanics (ILSB), Vienna University of Technology
* Correspondence: amir.horr@ait.ac.at; Tel.: +43-(0)50550-6918

Abstract

Artificial Intelligence (AI) and its subset, Machine Learning (ML), play transformative roles in the manufacturing sector, forming the foundation of the "Industry 4.0 and 5.0" frameworks. This research contributes to that evolution by developing AI-based advisory systems that utilize advanced data models to optimize casting processes. These systems exemplify the principles of smart manufacturing, where machines and processes are interconnected, adaptive, and driven by data. They support key objectives such as automation, seamless connectivity, real-time data exchange, human-centric innovation, operational resilience, and sustainability. The models developed in this work enable manufacturers to fine-tune product quality, minimize waste, and accelerate time-to-market through predictive analytics and dynamic process control. By integrating AI-based advisory systems, hybrid modeling, and reduced-order modeling techniques, the systems facilitate real-time decision-making and continuous improvement—essential for achieving flexible, efficient, and customized production environments. A real-world case study further demonstrates the effectiveness of these AI-based advisory systems in casting applications, detailing the steps involved in database construction, data training, and predictive modeling.

Keywords: data models; machine learning; neural network; manufacturing processes; digitalization; digital advisory system

---

1. Introduction

The integration of AI into manufacturing processes, particularly within the casting industry, marks a transformative milestone in the evolution of industrial practices under the paradigms of Industry 4.0 and Industry 5.0. Casting, as a fundamental manufacturing process involving metal melting, pouring, and solidification, is inherently complex due to the multi-physical 1hermos-fluid and metallurgical phenomena that govern product quality. Variations in process parameters such as melt temperature, casting speed, cooling rate, and mold geometry often lead to defects including cold cracking, hot tearing, porosity, shrinkage cavities, or residual stresses, which in turn drive up costs through rework, scrap, and delays. Traditional optimization approaches, while valuable, are often limited by their reliance on limited lab-scale trial-and-error experimentation, offline simulations, and operator intuition, which fail to meet the demands of modern, adaptive, and highly competitive manufacturing environments [1–4].

AI-based advisory systems address these challenges by combining data generation, machine learning, and predictive modelling to enable optimization of casting processes. Drawing on data from limited experimental campaigns, high-fidelity physics-based simulations, and sensors, these systems can capture the nonlinear, multi-scale nature (e.g., micro-macro microstructure evolution) of casting dynamics and translate it into efficient predictive framework. By doing so, they support predictive design, online quality control, and informed decision-making during production. Surrogate modelling and reduced-order modelling approaches, as advanced in recent research works [5–9] are

 doi:10.20944/preprints202604.0566.v1

particularly critical to this effort because they can generate general synthetic mass data and transform computationally expensive simulations into fast, real-time models suitable for integration into industrial advisory frameworks.

Furthermore, digital physics-based advisory systems for manufacturing have evolved into advanced decision-support—and increasingly decision-making—frameworks that integrate ML, statistical modeling, optimization techniques, and physics-based simulations. These systems aim to recommend, or autonomously implement, process set-points that optimize key performance indicators such as product quality, cost efficiency, throughput, and sustainability [10–13]. Over the past decade, their capabilities have progressed from offline analytics and dashboard-based monitoring to closed-loop, real-time control architectures embedded within digital twins, enabled by IoT-driven data pipelines and high-performance computing infrastructures. Comprehensive reviews across the manufacturing domain highlight a paradigm shift from descriptive analytics toward predictive and prescriptive intelligence, while also identifying persistent challenges related to data availability, quality, interoperability, and interpretability in industrial-scale deployments. Increasingly, these systems exhibit both direct and inverse prediction capabilities, supporting the design, monitoring, and optimization of complex manufacturing processes characterized by multi-physical, multiphase, and multiscale phenomena, as well as large temporal and spatial data domains that pose significant challenges to traditional modeling approaches [14–16].

This research investigates the deployment of digital advisory systems for modeling and designing material processing under industrial conditions. It systematically ad-dresses critical challenges, including data availability and integrity, the selection of suitable numerical solvers and interpolation schemes, snapshot generation for model reduction, validation protocols, and the development of generative data models and structured databases. The novelty of this work lies in the integrated framework that combines ML algorithms, physics-based solvers, and advanced interpolation techniques to manage both steady-state and transient manufacturing scenarios. Leveraging reduced-order modeling and iterative data-driven training, the proposed approach achieves substantial reductions in computational cost while enhancing predictive accuracy—particularly in environments characterized by multi-physical, multiphase, and multiscale complexity. Real-world case studies are presented to demonstrate the practical implementation and effectiveness of these concepts. Additionally, the study emphasizes ongoing efforts in process data acquisition and synthetic data generation, providing guidance on best practices for solver and interpolator selection, as well as practical considerations for industrial deployment.

2. Background and Literature Review

The integration of AI-driven data science techniques into manufacturing processes traces its origins to the early 20th century, beginning with the application of fundamental statistical methods for process control. This evolution has accelerated into a highly sophisticated manufacturing paradigm where data informs product design (e.g., generative models for concept development), process optimization (e.g., predictive algorithms for resource allocation), and process control (e.g., adaptive feedback mechanisms for operational stability). This section provides a chronological overview of this progression, highlighting key milestones, challenges, and technological breakthroughs. The historical context is rooted in the automation and computerization introduced during the Industry 3.0 era, where programmable logic controllers (PLCs) and early robotics established the foundation for advanced digital systems. Prior to this, the origins of data science in manufacturing can be traced to statistical process control (SPC), which introduced quantitative methodologies for monitoring and reducing variability in production processes [17–19].

In 1924, researchers introduced the first control chart—a graphical tool designed to distinguish common-cause from special-cause variations—establishing the foundation for data-driven decision-making in process control [20–22]. Subsequent advancements included the development of management principles such as Deming's 14 Points, the integration of statistical analytics into quality control frameworks, and the emergence of Total Quality Management (TQM) [23,24]. Early

 doi:10.20944/preprints202604.0566.v1

applications primarily focused on design and optimization through sampling-based techniques, which formalized data analytics for cost-effective process adjustments. The 1950s marked the onset of the computational revolution and the formal inception of AI, with profound implications for manufacturing automation. Following the introduction of the term "artificial intelligence," research began exploring machines capable of simulating human reasoning for design and control tasks. Early developments included neural networks (NNs) demonstrating pattern recognition for quality inspection and machine learning (ML) techniques applied to defect detection in assembly lines. By the 1960s, numerical control (NC) systems integrated basic data analytics into machine tools, employing punched cards for precise design replication—paving the way for computer-aided design (CAD). Experimental robotic systems soon emerged, incorporating rule-based decision logic and combining statistical models with elementary optimization algorithms [25–28].

The 1970s witnessed the resurgence of expert systems (ES), introducing rule-based AI designed to emulate human expertise for decision support in complex environments such as chemical analysis, which significantly influenced manufacturing ecosystems. These systems optimized tasks such as computer configuration design using if–then rules derived from engineering knowledge, reducing errors by up to 95% and saving millions in manufacturing costs. Subsequently, fuzzy logic was incorporated into ES to manage uncertainty in optimization processes, notably in fault diagnosis for chemical plants. The second major wave of AI integration occurred in the 1980s, driven by the application of ES to robotics and production scheduling. This era saw the emergence of tools enabling declarative programming for advanced production planning, marking a transition toward more intelligent and adaptive manufacturing systems [29–31].

The advent of Industry 4.0, initiated around 2011 in Germany, marked a transformative era in manufacturing through the introduction of "smart factories." This paradigm integrates cyber-physical systems (CPS), the Internet of Things (IoT), and cloud computing, with AI and ML serving as core enablers. These technologies facilitate real-time data acquisition and analytics, predictive maintenance, and autonomous decision-making, fundamentally reshaping the design, optimization, and control of manufacturing processes. By leveraging interconnected systems and advanced computational intelligence, Industry 4.0 establishes a data-centric ecosystem that drives operational efficiency, flexibility, and resilience across global production networks [32–39].

The development of surrogate models for engineering design and process optimization fundamentally relies on the rapid and accurate solution of parametric partial differential equations (PPDEs), in which the parameters may describe geometric configurations, material properties, boundary conditions, or processing variables. Depending on the methodological framework employed, surrogate modeling strategies can generally be classified into three main categories [40]. The first category comprises dimensional or analytical model reduction techniques, where the complexity of the governing equations is reduced through physical insight, for example, by exploiting symmetry to reduce three-dimensional formulations to two- or quasi-two-dimensional problems, or by replacing nonlinear systems with linearized approximations. The second category includes computational model-order reduction methods, such as the reduced basis (RB) method and proper orthogonal decomposition (POD), which construct low-dimensional representations of high-fidelity numerical models through projection onto dominant modal subspaces. The third category consists of data-driven and physics-informed machine learning approaches, including physics-informed neural networks (PINNs) and deep operator networks (DeepONets), which leverage modern learning architectures to approximate parametric solution operators directly.

Within this landscape, physics-informed machine learning has attracted increasing attention due to its ability to combine the expressive power of neural networks with the rigor of physical modeling. In PINNs represent a natural and systematic integration of machine learning and scientific computing [41–43]. Although the conceptual foundations of embedding differential equations into neural network training date back to pioneering works in the 1990s, the modern formulation introduced by Raissi, Perdikaris, and Karniadakis in 2017 [44] significantly advanced the methodology and established its widespread adoption. This physics-based regularization constrains the hypothesis

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0566.v1

space and enables the learning of physically consistent solutions even in scenarios with sparse or partially missing data.

In manufacturing applications, physics-informed approaches have demonstrated promising capabilities for modeling complex process phenomena. For example, PINNs have been employed to predict solidification directions in freeze casting processes, to analyze curing and material behavior in composites manufacturing, and to infer coating conditions in blade coating operations [45]. These applications highlight the suitability of physics-informed machine learning for manufacturing environments, where governing equations are known but full-field measurements are limited, and where rapid prediction under varying process parameters is essential [46]. By embedding physical structure into data-driven frameworks, such methods offer a compelling pathway toward accurate, computationally efficient, and physically interpretable surrogate models for advanced manufacturing systems [47]. Figure 1 illustrates the pictorial representation of data methods background over a century from statistical process control to today's Industry 4.0 and 5.0 smart factories.

Figure 1. Pictorial representation of data methods background over a century from statistical process control to today's Industry 4.0 and 5.0 smart factories.

3. Data Processing and AI Applications

The evolution of data processing and AI applications in manufacturing represents a transformative progression from basic statistical methods to advanced cyber-physical systems, including full digital twins, digital shadows, and AI-driven advisory systems. As computational power and data availability increased, the emergence of data mining and intelligent analytics enabled industries to identify hidden patterns and inefficiencies in production processes, thereby enhancing decision-making. Today, the integration of AI techniques with cloud computing and edge-based analytical capabilities supports real-time monitoring and adaptive control. The latest phase of this evolution is defined by digital twins, digital shadows, and AI-enabled advisory systems, which provide dynamic virtual representations of physical assets that are continuously synchronized with real-world data.

3.1. Digital Systems for Manufacturing

The methodological framework for digital systems in manufacturing is based on a multi-stage approach that integrates data acquisition, processing, modeling, and decision support.

- At its foundation, data acquisition is achieved through data mining, experimental trials, numerical simulations, and real-time sensor streams, ensuring comprehensive visibility into process characteristics and production parameters.
- The collected raw data is then subjected to filtering, mapping, translation, interpretation, and preprocessing to improve quality, consistency, and suitability for advanced analytics.
- The subsequent stage focuses on data modeling and analytics, where dimensionality-reduction techniques, statistical methods, and machine learning algorithms—such as regression, clustering, and neural networks—are applied to optimize processes, detect anomalies, and identify faults.
- Building on these physics-informed and/or physics-based models, real-time data models, simulation environments, and virtualization frameworks enable the development of digital twins, digital shadows, and digital advisory systems. These technologies create dynamic, high-fidelity virtual representations of physical assets and manufacturing processes. These virtual models are continuously synchronized with real-world data, allowing for scenario evaluation, performance prediction, and optimization under varying operational conditions. Finally, AI-driven advisory and autonomous systems leverage advanced decision engines to deliver prescriptive recommendations or execute automated control actions, forming the foundation of smart manufacturing ecosystems.

The completed framework emphasizes interoperability, scalability, and cyber-physical integration, enabling manufacturers to transition from reactive operations to predictive and prescriptive strategies, ultimately enhancing resilience, operational efficiency, and sustainability. Figure 2 illustrates the pictorial representation of methodological framework for digital systems in manufacturing.

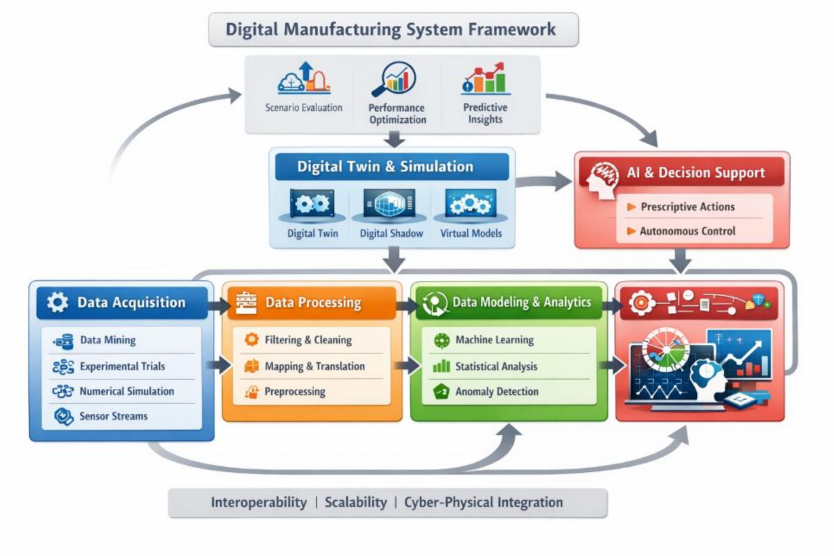


Figure 2. Pictorial representation of methodological framework for digital systems in manufacturing including data acquisition, processing, modeling, and decision support.

3.2. Data Sources & Databases

A robust framework for building data sources and production databases for manufacturing processes such as casting, extrusion, and additive manufacturing (AM) begins with structured, interoperable process data and progresses through semantic data models. The inherent multi-physical, multi-scale, and multi-phase nature of these processes, combined with their complex parameter spaces, presents significant challenges in generating high-quality data for database construction.

Experimental Data: The most trusted source of process data is experimental data, obtained through direct physical measurements and observations. These datasets provide high fidelity but are often costly and time-consuming to acquire. For processes with large and complex parameter spaces, the number of required experimental trials can be prohibitively high, making this approach expensive. Consequently, experimental trials are typically limited to validation purposes, while the main body of process databases is constructed using complementary methods.

Numerical Simulation Data: Validated numerical simulations represent one of the most efficient means of generating process data. By defining numerical domains with varying initial and boundary conditions across the parameter space, simulations can replicate real-world scenarios at scale. With advances in computational power and modeling techniques, simulation-based data generation has become a cornerstone for building comprehensive process databases.

Mined Data: For processes with relatively small parameter spaces, mining data from existing literature and publicly available databases can be an effective strategy. This approach leverages prior research and industrial benchmarks to supplement experimental and simulated datasets.

Live Process Data: Sensor-based data acquisition and monitoring systems provide dynamic, real-time information from operating equipment. These live data streams are essential for constructing feedback loops and adaptive models, enabling continuous improvement of predictions and process control through iterative updates. For complex, multi-physical, and multi-scale processes such as casting, extrusion, and AM, the initial process database should combine limited experimental trial data, verified numerical simulation outputs, and, where applicable, mined data from external sources. Over time, live process data can enhance and refine these databases through feedback mechanisms, improving accuracy, robustness, and adaptability for advanced applications such as digital twins and AI-driven advisory systems. Figure *3* illustrates the pictorial representation of methods for data sourcing and process database building.

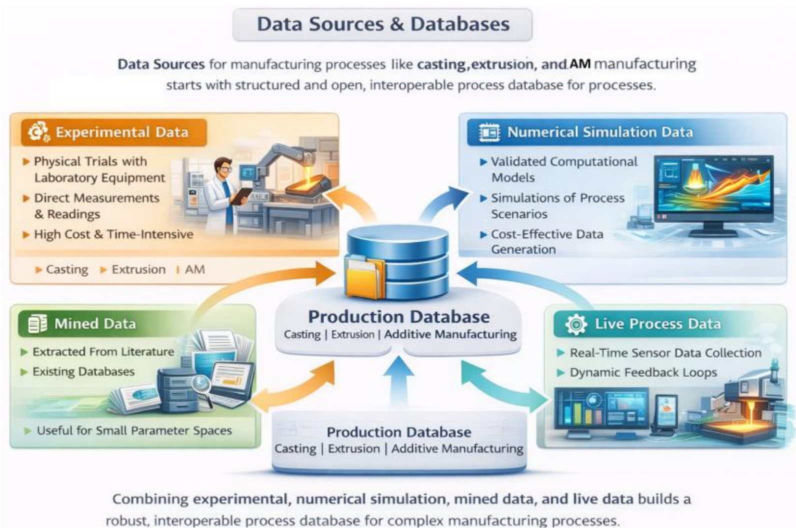


Figure 3. Pictorial representation of methods for data sourcing and process database building using experimental, simulations, mined and live data.

3.3. Reduced-Order Models

Reduced order models (ROMs) in manufacturing processes are commonly defined as systematic modeling approaches for deriving low-dimensional, computationally efficient surrogate models that accurately approximate the dominant physics and process behavior of complex manufacturing systems. These models originate from high-fidelity, multi-physics, multi-phase, and multi-scale simulations that are typically based on physics-informed or fully physics-based formulations. In this context, ROMs aim to retain the essential characteristics of manufacturing processes, including dynamic behavior, thermomechanical interactions, phase transformations, microstructural evolution,

 doi:10.20944/preprints202604.0566.v1

and nonlinear process responses—while substantially reducing the computational cost compared to full-order numerical simulations such as finite element or finite volume models. Key physical phenomena represented by ROMs may include heat transfer, fluid flow, and material deformation.

ROM techniques applied to manufacturing are commonly grounded in rigorous data science and mathematical methods, including proper orthogonal decomposition (POD), reduced basis methods, surrogate regression models, and machine-learning-assisted ROMs. These approaches are particularly well suited for real-time or near–real-time prediction of manufacturing processes such as casting, extrusion, and additive manufacturing, where the underlying parameter spaces are high-dimensional and rapid decision-making is essential. In the context of casting processes examined in this research, a range of data-solver techniques can be employed to disentangle the multi-dimensional parameter space and construct efficient reduced models. These techniques include eigenvalue-based methods, regression approaches, and classification algorithms. To enable real-time prediction, the data solvers are coupled with efficient data interpolation techniques, most commonly radial basis functions (RBF) and inverse distance weighting (IDW). For enhanced data training and pattern recognition, machine learning methods—such as neural networks (NNs) and genetic-algorithm-based symbolic regression—can be integrated into the modeling framework to improve predictive capability and robustness.

The development of reduced models for manufacturing processes has been ad-dressed in previous studies [48–52], where integrated frameworks combining data solvers, interpolation strategies, and machine learning routines were established. In this work, an eigenvalue-based data solver is developed for a multi-physics, multi-scale casting application, employing singular value decomposition (SVD) to decompose, project, and reduce the dimensionality of the process data. Four key casting parameters are considered: casting speed, initial melt temperature, cooling water flow rate, and cooling water temperature. Together, these parameters define a four-dimensional process space for the reduced order model.

Two sets of reduced models are generated in that study. First, a small database consisting of 16 casting scenarios is constructed to develop an initial reduced model for verification and validation purposes. These scenarios are generated using Latin hypercube sampling (LHS) to ensure a well-distributed coverage of the parameter space. Following successful verification and validation, a larger database comprising 100 casting cases is used to construct a second set of reduced models for large-scale data generation. The same sampling strategy is applied to populate the expanded database, and the resulting reduced models are then used to generate 10,000 new predictions of thermal and microstructural fields. These predictions serve as training data for the process advisory framework. Figure 4 illustrates the schematic overview of the reduced order modelling framework, while, Figure 5 shows performance of the data models for thermal and microstructure predictions, where selected computational fluid dynamics (CFD) design-of-experiments (DOE) cases are used for model validations.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0566.v1

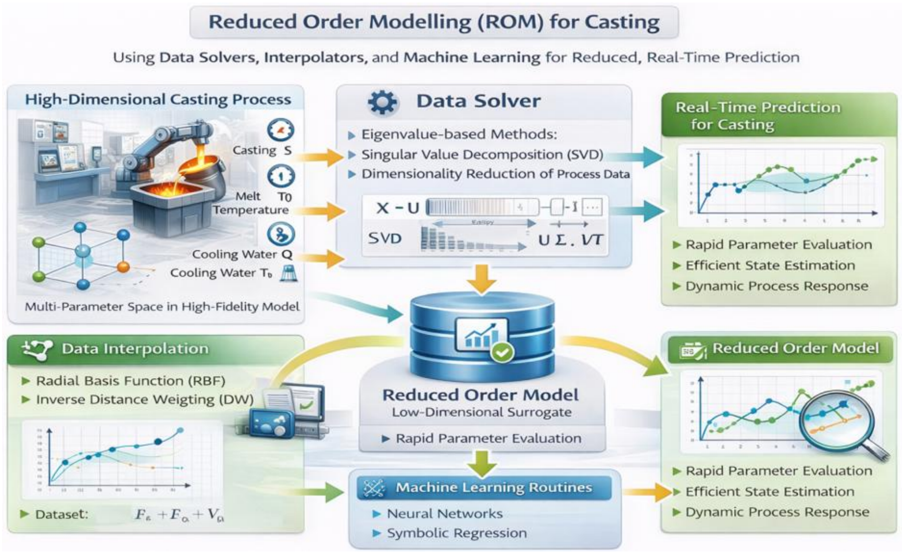


Figure 4. Schematic overview of reduced order modelling framework including its components data solvers, interpolators and parameter space.

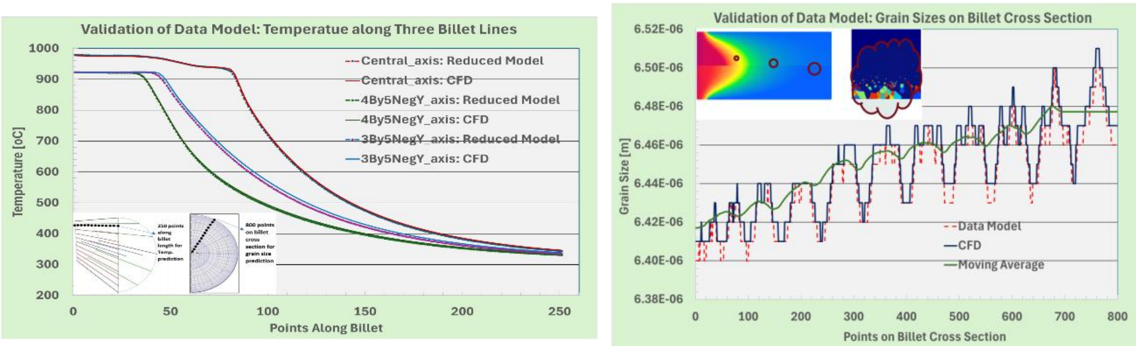


Figure 5. Data models performances for thermal and microstructure predictions against CFD simulations for some DOE cases.

4. AI-Based Advisory System

In the context of manufacturing digitalization, digital advisory systems—often referred to as online process advisory systems—are best understood as AI-enabled design and decision-support tools. These systems can operate atop a process's data infrastructure and its virtual counterparts (such as digital shadows and digital twins), providing guidance for new design scenarios and continuously transforming available process data into real-time predictions and corrective recommendations for engineers and operators. As an AI-powered platform, it leverages real-time predictive models to support process design and execution. It identifies potential failures and quality deviations, recommending optimal parameter sets while enhancing (rather than replacing) human expertise in the manufacturing environment.

4.1. System Architecture

Many industrial manufacturing processes are governed by a set of operational parameters that must be carefully selected, controlled, and continuously monitored throughout production. The appropriate choice of these parameters is critical, as certain combinations may introduce safety hazards, cause unplanned downtime, or compromise product quality. Identifying parameter configurations that ensure stable and reliable operation while simultaneously meeting stringent quality requirements is a complex multidimensional task. In industrial practice, this process often involves iterative trial-and-error procedures and depends heavily on the expertise and tacit

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0566.v1

knowledge of experienced operators. Moreover, even minor deviations from optimal operating conditions can result in quality deterioration, reduced process efficiency, or production interruptions.

An AI-based advisory system is designed to assist operators by providing data-driven recommendations for feasible and robust process settings under varying operational constraints. In real-world manufacturing environments, not all process parameters can be adjusted freely: some variables can be modified dynamically during operation, whereas others may be temporarily fixed due to physical, technical, or plant-specific limitations. The advisory task can therefore be formulated as a constrained inverse problem: given a scenario in which a subset of process parameters is fixed, determine appropriate values for the remaining adjustable parameters such that the process operates safely and stably, and the resulting product satisfies predefined quality criteria. By systematically exploring feasible regions of the parameter space, such systems aim to enhance decision-making, reduce reliance on trial-and-error experimentation, and improve overall process robustness and efficiency.

As illustrated in Figure 6, operators first provide the current process state and specify which parameters are fixed. These fixed inputs define the operational constraints within which the AI-based advisory system must identify feasible adjustments. Based on this constrained scenario, a generative model proposes candidate values for the remaining adjustable parameters. Each candidate configuration is subsequently evaluated according to criteria developed in collaboration with domain experts. These criteria quantify process stability, product quality, and any additional process-specific requirements. Configurations that fail to meet the predefined minimum thresholds for any criterion are systematically discarded. The remaining feasible candidates are then ranked using a composite, process-specific performance metric—often referred to as a fitness score—that integrates all relevant quantitative measures into a single evaluation framework. The highest-ranking recommendation(s) are ultimately presented to the operators via a dedicated Human–Machine Interface (HMI). With appropriately selected AI architecture and a suitable computational infrastructure, this advisory workflow can be executed within a few seconds, thereby enabling near real-time decision support during production. While the described forward (online) workflow illustrates the operational use of the system, its overall performance critically depends on the underlying AI model and the robustness of its offline training pipeline. In particular, model selection, data quality, and feature representation significantly influence predictive accuracy and generalization capability. This work focuses on generative AI models and deep learning techniques, which typically require large-scale, high-quality datasets and carefully designed data representations to ensure reliable performance in industrial environments.

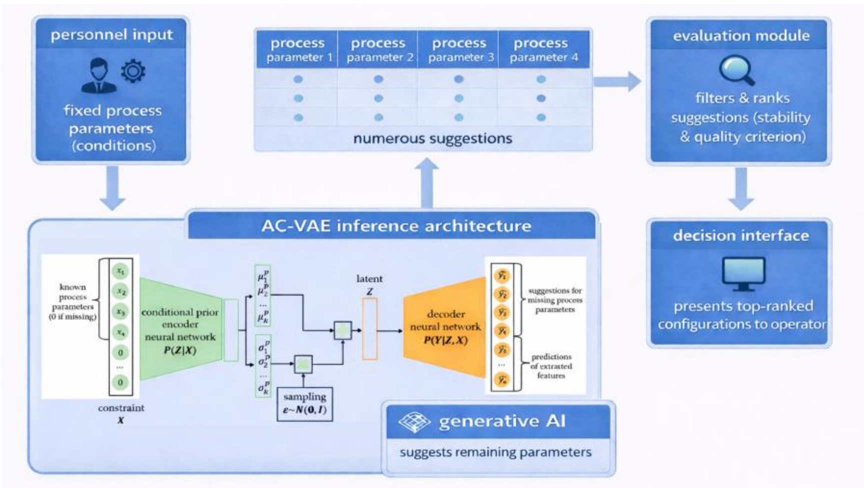


Figure 6. Schematic overview of the proposed AI-based advisory system and its online decision-support workflow.

4.2. Data Infrastructure, Generation & Flow

Ideally, the data acquisition stage should rely on real physical measurements obtained directly from the manufacturing process. In industrial practice, however, such data are often costly to collect, limited in quantity, and constrained by operational, safety, and production requirements. These limitations restrict systematic exploration of the parameter space and hinder the development of data-intensive AI models. To address this challenge, the present work relies on synthetic data generated through high-fidelity numerical process simulations or data-driven surrogate models derived from them. While high-fidelity simulations provide accurate representations of the underlying physical phenomena, their computational cost can be prohibitive for large-scale data generation. Therefore, complementary strategies for efficient mass data generation are introduced to achieve broad and systematic coverage of the relevant parameter space while maintaining physical consistency.

To enable a robust simulation-driven workflow, a dedicated data infrastructure was developed to ensure reproducibility, traceability, and efficient data management. Process-relevant information is stored in a PostgreSQL database, serving as the central repository for simulation inputs and post-processed outputs. The numerical simulation pipeline is deployed in a cluster computing environment, and simulation runs are orchestrated through automated scripts managing data exchange between the database and the simulation framework. The database schema follows a relational model, whose complexity depends on the number of modeled variables and the level of detail in the simulated process. A central component of the schema is the result table, storing post-processed simulation outputs together with foreign-key references to associated metadata tables, including process parameter combinations, material properties of the modeled system, geometric or mesh-related inputs, and other process- or simulation-specific descriptors. This structured organization ensures full traceability: for any given simulation result, all underlying inputs and configuration details can be retrieved through straightforward database queries, while reproducibility is ensured by persistently storing the complete set of metadata required to rerun a simulation under identical conditions. Consequently, the proposed infrastructure provides a scalable foundation for large-scale synthetic data generation and AI model development in manufacturing applications.

Post-processing constitutes a critical stage, as it determines which components of the raw simulation output are preserved and how they are represented within the database. High-fidelity numerical simulations typically generate three-dimensional fields of multiple physical variables evolving over time on fine-resolution meshes. Depending on the characteristics of the manufacturing process and the definition of the advisory task, either the full transient solution must be retained, or only steady-state information is required. In the latter case, storing solely the final time step is sufficient, and intermediate states can be omitted. Persisting full 3D mesh data for every simulation, particularly for each time step is highly demanding in terms of storage capacity and increases downstream processing complexity. In practice, ML models rarely require complete high-resolution 3D fields, as advisory tasks typically depend on characteristic patterns, physically meaningful summary metrics, or engineered features capturing essential process behavior. Consequently, a central objective of post-processing is the transformation of high-dimensional mesh-based outputs into compact, information-rich representations that preserve the relevant physical content while eliminating redundant spatial detail. This dimensionality reduction step reduces storage requirements, improves computational efficiency, and facilitates model training and inference. During the exploratory development phase, multiple data representation strategies can be evaluated based on compressed versions of the full mesh stored in the database. These may include one-dimensional engineered feature vectors, composed of statistical or physically motivated descriptors such as spatial averages, extrema, gradients, or other characteristic quantities; two-dimensional image-like representations constructed by extracting strategically chosen slices of the three-dimensional; or reduced three-dimensional fields generated by coarsening the original mesh. Multi-modal representations are also possible, combining tabular process parameters with image-based or reduced-field inputs to exploit complementary information sources. Each representation entails a trade-off between retained physical detail and computational cost.

The overarching objective is to identify the simplest representation that enables the advisory system to achieve high predictive performance while meeting real-time inference requirements. Following this exploratory phase, the selected data format can be fixed and the post-processing pipeline streamlined accordingly, ensuring that only the chosen representation is stored in the database. This consolidation reduces data management complexity, enhances scalability, and allows subsequent efforts to focus on expanding the dataset and refining the AI model architecture.

4.3. Mass Data Generation

This research introduces a structured, multi-stage data modeling methodology specifically designed to address a major bottleneck in smart manufacturing: the scarcity of high-quality, high-volume, and well-distributed process data. The proposed approach systematically transforms a limited, high-fidelity dataset—obtained from experiments and calibrated CFD simulations—into a large-scale synthetic database comprising more than 10,000 process scenarios. This expanded dataset is subsequently used for training data-intensive machine learning models within an online, real-time process advisory system.

The methodology begins with the construction of initial ROMs, developed using eigenvalue-based solvers and IDW interpolation techniques. These preliminary data models are calibrated and validated against high-fidelity CFD simulations. Validation is performed using additional DOE scenarios that deliberately explore normal operating conditions, near-limit regimes, and beyond-limit cases to ensure robustness and physical consistency. Following comprehensive verification, an extended sampling strategy is applied to generate a larger snapshot matrix comprising approximately 100 casting scenarios. These CFD simulations form an enriched response database covering the multidimensional parameter space, including variables such as melt temperature, cooling conditions, and casting speed.

Building upon this validated response database, a scalable framework for mass synthetic data generation is established. A second-generation data model is constructed using SVD-based solvers in combination with IDW interpolation, enabling efficient parametric mapping of process responses. Subsequently, a large-scale snapshot matrix containing more than 10,000 parameter combinations is generated, and synthetic process responses are rapidly computed using the ROM instead of performing additional CFD simulations. Because the ROM retains the dominant eigenmodes of the governing physical system, predictions within the calibrated parameter range remain physically coherent, thermally consistent, and smoothly varying across the parameter space—while requiring only a fraction of the computational cost of full-order simulations.

The resulting ROM-generated synthetic dataset preserves the structural patterns, correlations, and physical constraints embedded in the calibrated high-fidelity simulations. This dataset is then integrated into the training pipeline of deep learning models, including neural network-based architectures, which typically require thousands of representative samples for stable and accurate learning. By leveraging ROM-based mass data generation, the methodology eliminates the prohibitive computational time, cost, and experimental effort associated with large-scale CFD simulations or physical trials.

Figure 7 illustrates the complete workflow for large-scale synthetic data generation and its integration into the AI training framework for online advisory applications.

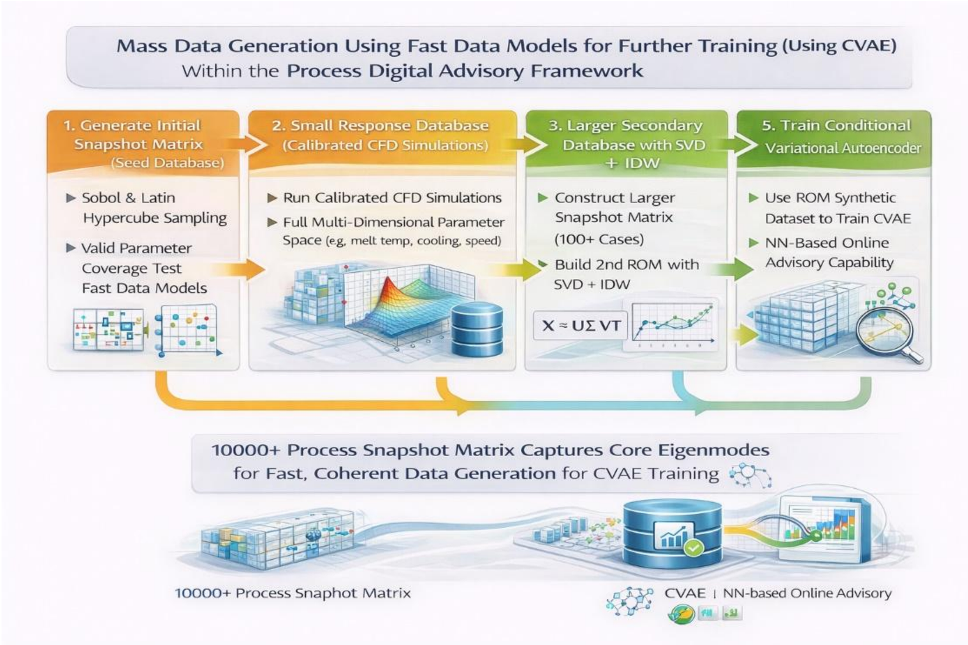


Figure 7. Workflow for the generation of mass synthetic data for further data training using the data model.

4.4. Neural Network Learning

The choice of neural network architecture depends on the data representation. For one-dimensional tabular features, feed-forward neural networks (NNs) are typically appropriate; for two-dimensional image-based representations, convolutional neural networks (CNNs) are preferred; and for three-dimensional or irregularly structured data, graph-based or other specialized architectures often provide superior performance. The present work focuses on tabular feature representations, while also investigating the potential benefits of a multimodal framework incorporating complementary image-based information.

Within the tabular setting, each simulated process scenario is encoded as a feature vector of dimension n, comprising quantitative descriptors of process parameters and key physical quantities derived from simulation outputs. Under this formulation, the advisory task can be interpreted as a feature imputation problem: given a subset of known features corresponding to fixed process parameters, the objective is to infer the remaining features, including adjustable process parameters and relevant physical response variables. To address this imputation task, generative AI models provide a principled framework. In this study, we focus on Variational Autoencoder (VAE)-based architectures [53]. VAEs are probabilistic generative models that learn a structured latent space distribution from which realistic data samples can be generated. The Conditional Variational Autoencoder (CVAE) extends this framework by incorporating conditioning variables that guide the generative process; detailed mathematical derivation of both VAEs and CVAEs are provided in [54]. In the context of feature imputation, CVAEs can reconstruct missing or masked components of the input vector by leveraging the learned latent representation while conditioning on the available features. However, the standard CVAE formulation assumes a fixed conditioning structure defined during training.

In real industrial advisory scenarios, however, this assumption is overly restrictive. Different operating situations may involve different subsets of fixed parameters; in principle, any subset of process variables may be observed or unknown at inference time. To address this limitation, this work adopts the Variational Autoencoder with Arbitrary Conditioning (AC-VAE) [55], which is trained to handle arbitrary subsets of observed features and infer the remaining components. From a probabilistic standpoint, this flexibility arises from a crucial modification of the conditional prior distribution. In a standard CVAE, the latent variable Z is assumed to follow a standard normal prior

 doi:10.20944/preprints202604.0566.v1

independent of the conditioning variables. In contrast, AC-VAE models the conditional prior $P(Z|X)$ as a Gaussian distribution

$$P(Z|X) = \mathcal{N}\left(\mu_P(X), \Sigma_P(X)\right), \quad (1)$$

where the mean $\mu_P(X)$ and covariance $\Sigma_P(X)$ are learned functions of the available conditioning variables X. Architecturally, while a CVAE consists of two NNs—the encoder and the decoder—an AC-VAE employs three networks: (i) the proposal encoder Q (ii) the decoder P; and (iii) a conditional prior encoder approximating $P(Z|X)$. The complete training architecture is illustrated in Figure 8.

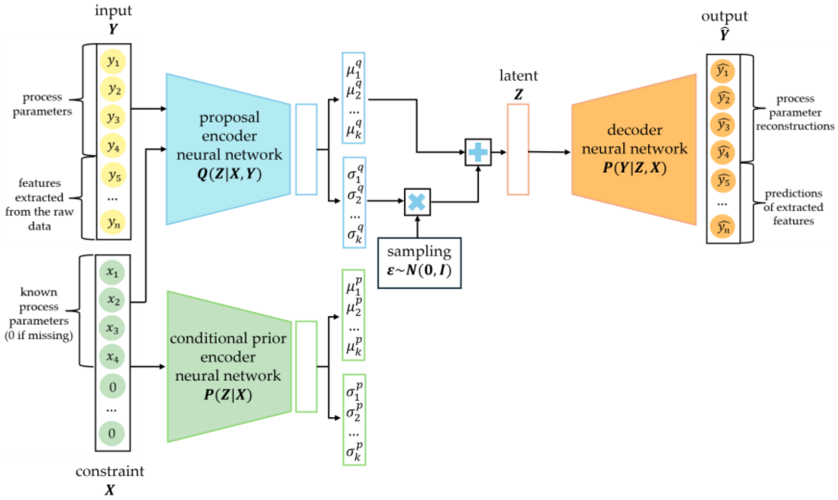


Figure 8. Training-time architecture of the AC-VAE model.

The training of NNs has become highly accessible through modern high-level libraries. In this work, the custom AC-VAE architecture was implemented in PyTorch, selected for its flexibility and suitability for non-standard models. To reduce boilerplate code and streamline experimentation, PyTorch Lightning was employed, enabling structured training workflows and seamless integration of advanced features such as multi-GPU training, experiment logging, and automated hyperparameter optimization. Optimization is based on gradient descent, enabled by the differentiability of the network architecture and its loss function. The objective combines two terms: a Mean Squared Error (MSE) term that measures reconstruction fidelity under conditioning constraints, and a Kullback–Leibler Divergence (KLD) term regularizing the latent space by aligning the approximate posterior with a conditional Gaussian prior. The closed-form expression of the KLD under Gaussian assumptions ensures stable and efficient training. The Adam optimizer was chosen for its robustness and adaptive gradient handling, and additional regularization techniques were explored to mitigate overfitting.

The dataset is partitioned into training, validation, and test sets using a 60:20:20 split, with Z-score standardization applied to harmonize feature scales. To handle partially observed inputs, each sample is paired with binary masks encoding observed and unknown parameters. The constraint vector is concatenated with its mask, enabling the model to distinguish masked entries from genuine zero-valued features. In the bimodal setting, cross-sectional mesh data are interpolated into 2D image representations and processed through a U-Net-inspired convolutional encoder [56] without skip connections. Extracted image features are fused with tabular features and passed to the AC-VAE. Reconstruction is performed symmetrically, with an additional pixel-wise MSE loss incorporated to enable end-to-end optimization of the combined CNN–NN architecture.

At inference time, the proposal encoder is discarded, and only the conditional prior encoder and decoder are retained for generation (Figure 6). While standard regression metrics can be computed on test data, evaluating true generative capacity remains challenging. Although generated samples satisfy imposed constraints and are drawn from the learned data distribution, their physical consistency cannot be guaranteed without coupling to external simulations (e.g., a ROM). Consequently, generative performance is assessed on fully observed test cases using regression metrics, with the coefficient of determination (R^2) preferred for its normalized and interpretable scale.

5. Case Study: Casting Application

To demonstrate the practical applicability of the proposed digitalization framework, a representative casting use case is introduced. This demonstrator focuses on steady-state casting process conditions, deliberately excluding transient effects associated with process start-up in order to ensure a controlled and reproducible evaluation environment. By concentrating on stable operating regimes, the case study enables a clear assessment of how data-driven models can support process design and decision-making under realistic production conditions. The implementation integrates the essential components of a modern digital workflow, including systematic data acquisition, preprocessing, model development, and training based on an AC-VAE. Within this framework, an online advisory system is realized that provides real-time insights and actionable recommendations.

5.1. Experimental & Simulation Setup

Horizontal Direct Chill (HDC) casting is a continuous industrial process widely employed for the production of high-grade aluminum billets. In this process, molten aluminum undergoes initial heat extraction while flowing through the mold (primary cooling). Upon exiting the mold, the majority of the remaining thermal energy is removed by a circumferential jet of chilled water (secondary cooling). The high cooling rates achieved in HDC casting are critical for establishing the microstructural characteristics and mechanical properties required for subsequent downstream operations, most notably extrusion.

Stable operation of the HDC process depends on precise regulation of key process inputs. Persistent deviations from nominal operating conditions can rapidly impair billet quality, lead to unplanned production interruptions, and potentially introduce safety hazards. In the present study, billets of A6082 aluminum alloy were cast under industrial conditions. Process parameters were actively controlled until nominal steady-state conditions were attained. The internal thermal evolution of the billet was monitored using thermocouples inserted into the melt at predefined locations within the mold. These sensors traveled with the advancing billet, enabling the acquisition of temperature histories along multiple axial trajectories. This measurement strategy further allowed reconstruction of the solidification sump geometry.

A high-fidelity numerical model of the process was developed within the OpenFOAM computational fluid dynamics (CFD) framework. The model employs an Eulerian thermo-fluid formulation, conceptually aligned with the approach proposed by Bennon [57] and further extended to incorporate non-equilibrium solidification [58] effects and microstructural growth kinetics [59]. Assuming lateral symmetry, the computational domain represents one half of the billet and is discretized using a finite-volume mesh. The mesh remains fixed throughout the simulation, with designated inlet and outlet boundaries at the axial extremities. Custom boundary conditions are implemented at the external walls to represent heat extraction by the mold (primary cooling) and by the chilled water jets (secondary cooling).

Key geometrical and operational inputs are parameterized to facilitate systematic database generation. Among these, the initial melt temperature, casting speed, chill-water flow rate, and chill-water temperature were selected for variation in order to quantify their influence on process outcomes. Once quasi-steady conditions are reached, the solution fields are time-averaged over multiple time steps. The results are subsequently post-processed to extract thermal histories along selected axial paths, together with predicted microstructural characteristics at the billet outlet plane. Figure 9 illustrates the overlap of the simulation domain with the actual HDC process, including its volumetric 3D discretization and melt flow streamlines mapped with thermal results.

doi:10.20944/preprints202604.0566.v1

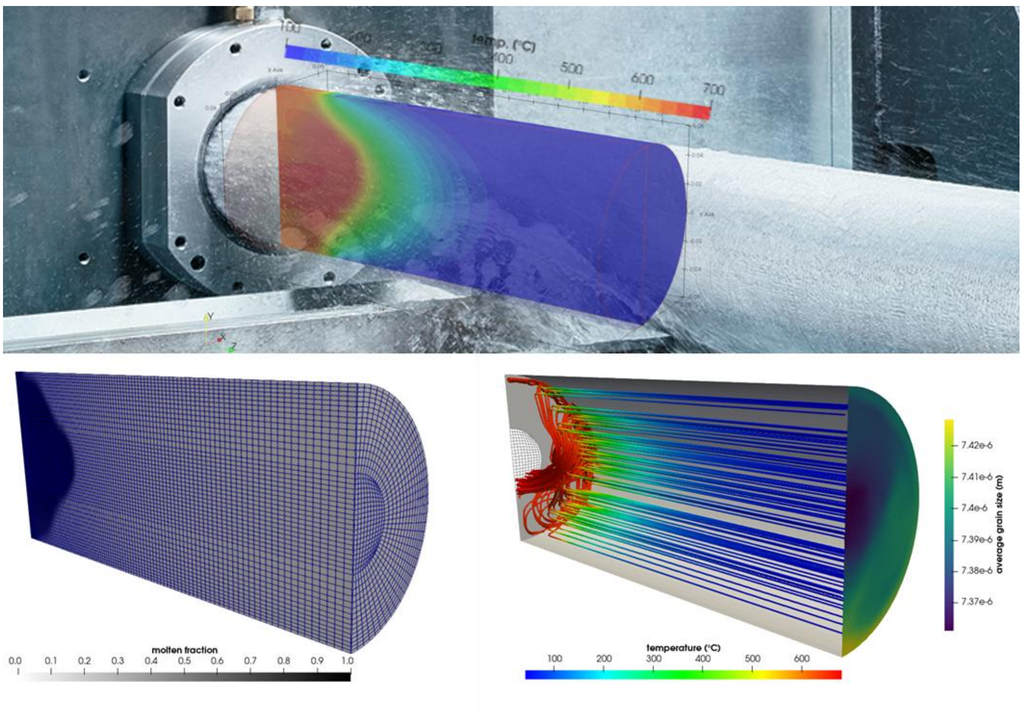


Figure 9. (top) Symbolic overlap of the simulation domain showing thermal results with the actual HDC process. (bottom-left) Volumetric domain discretization. (bottom-right) Melt flow streamlines mapped with thermal results, along with eventual average grain size results at the outlet domain boundary.

5.2. Database Building

The development of reliable process digital advisory systems depends fundamentally on constructing accurate, balanced, and well-covered databases that represent the full multidimensional parameter space of a manufacturing process. An effective strategy follows a tiered, progressive approach, beginning with a small, high-fidelity seed database for model validation and then expanding toward large synthetic databases generated through ROM-based process scenarios. The initial seed database is generated using Sobol or LHS techniques to ensure uniform, balanced, and non-clustered coverage of influential process parameters. For the present work on casting processes, these parameters include melt temperature, cooling rate, casting speed, water-cooling temperature, and other domain-specific variables. Each seed scenario is simulated using calibrated and experimentally validated CFD models, ensuring that the resulting response database faithfully captures the governing thermal and physical behavior of the process. This compact but well-distributed dataset provides the foundation for validating the initial data models, including their decomposition (e.g., POD/SVD) and interpolation strategies, which together form the mathematical backbone of the physics-informed data modeling framework.

Once the seed models demonstrate sufficient accuracy, a larger secondary snapshot matrix—typically comprising around 100 CFD-based process scenarios—is constructed to span the multidimensional parameter domain with greater resolution. This expanded database supports the development of more advanced surrogate and reduced-order models by capturing richer process variability while remaining computationally feasible. At this stage, ROM predictions are rigorously validated using independent DOE cases, including normal, near-boundary, and extrapolated operating conditions, ensuring robustness across the entire process space before scaling up further.

With a validated ROM established, the methodology advances to the generation of very large synthetic datasets (10,000+ scenarios). Using reduced-order models that en-code the dominant eigenmodes of the physical system, these synthetic datasets can be produced extremely rapidly (on the order of 900 seconds) while maintaining physical coherence, thermal consistency, and smooth behavior across parameter variations. Because the ROM inherits its structure from physics-validated

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0566.v1

CFD simulations, every predicted scenario remains grounded in the true physical constraints of the process, even when exploring large and dense parameter grids. These mass synthetic datasets then serve as the training foundation for the AC-VAE module within the digital advisory framework, enabling real-time predictive–corrective process design insights and intelligent online process-parameter recommendation capabilities.

5.3. Varitional Autoencoder with Arbitrary Conditioning

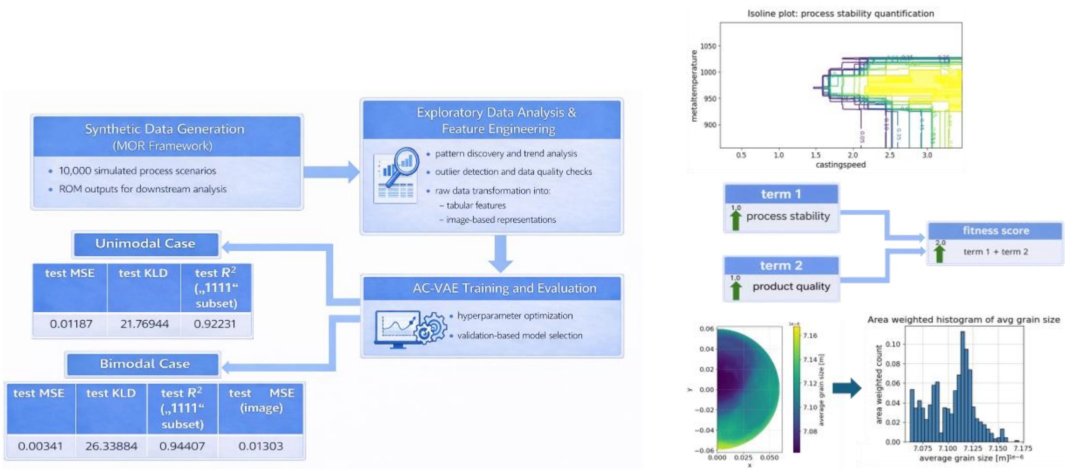
Extensive exploratory data analysis (EDA) was conducted on the ROM dataset to systematically assess the influence of process parameters on the raw output data. Based on these findings, feature engineering was performed to effectively represent each case as a high-dimensional feature vector. Derived features included melt pool depth, statistical descriptors of the grain size distribution at the outlet cross-section, temperature values sampled along the billet's central axis and surface at key locations, and multiple indicators related to liquid fraction evolution. To mitigate redundancy and multicollinearity, highly correlated variables were eliminated through feature selection, resulting in a compact representation of nine properties: four primary process parameters and five engineered features. For the bimodal configuration combining tabular and image-based data, the final billet cross-section was selected as the principal image representation due to its direct relevance to product quality. Scalar mesh values were projected onto a regular pixel grid, and linear interpolation computed pixel intensities as distance-weighted averages of neighboring mesh values, yielding smooth, physically consistent images. The inclusion of a longitudinal section as complementary image input remains a subject for further investigation.

The hyperparameter (HP) search space encompassed both architectural and training-related variables. Architectural hyperparameters included the number of hidden layers in the encoders and decoder of the AC-VAE, latent space dimensionality, and activation functions; in the bimodal setting, the convolutional encoder depth was also varied. Training-related hyperparameters—batch size, learning rate, weight decay, and the use of batch normalization—were systematically explored. The training objective combined tabular MSE and KLD (typically down-weighted), with an additional pixel-wise MSE term in the bimodal case; relative weights of these components were treated as hyperparameters and varied across trials. To ensure comparability between configurations, model selection was based exclusively on the tabular validation MSE. Early stopping was employed to mitigate overfitting. In total, 500 hyperparameter configurations were evaluated independently for the unimodal and bimodal settings, and the best models were selected based on validation loss.

The workflow summarized in Figure 10 outlines the principal stages of the proposed methodology. After completing hyperparameter optimization for both AC-VAE configurations, the best-performing configurations were retrained. Their performance was evaluated on an independent test set, with metrics reported in Figure 10. Both configurations demonstrate strong generalization capability, reflected by low reconstruction MSE values on the tabular features. Although the KLD is reported for completeness, its absolute magnitude is not directly comparable across models due to differences in latent space dimensionality. Nonetheless, the observed KLD values indicate an appropriate balance between latent space regularization and reconstruction fidelity. In particular, the magnitude of the KLD suggests active utilization of the latent variables, avoiding posterior collapse while maintaining high reconstruction accuracy.

A notable reduction in tabular reconstruction MSE is observed in the bimodal configuration, demonstrating that the inclusion of image-based information enhances tabular reconstruction. However, a fundamental limitation remains the absence of direct metrics for assessing generative capacity at inference time without coupling the model to online physical simulations. In the restricted scenario where all four process parameters are known (the "1111" subset in Figure 10 (a)), generative performance is evaluated using the R^2. The global R^2, averaged across features, shows a slight improvement for the bimodal model. More pronounced gains are observed on a per-feature basis, particularly for grain-size-related quantities, attributed to the explicit integration of grain-size information through the image modality during training. Despite these advantages, image reconstruction quality remains limited. Although the per-pixel MSE appears low, it is computed on scaled data and pixel-wise; qualitative inspection reveals noticeable reconstruction errors.

For integration into the proposed advisory system (Figure 6), the AC-VAE inference architecture generates multiple candidate parameter configurations under specified constraints. These suggestions are quantitatively evaluated with respect to process stability and product quality. A composite fitness score was defined to enable systematic filtering and ranking of generated candidates. Process stability is defined as the likelihood that a suggested parameter set produces acceptable outcomes while avoiding bleed-out and freeze conditions. While a deterministic stability assessment exists within the high-fidelity CFD framework, its reliability diminishes after the data reduction required for ROM construction, as full-field mesh information is no longer retained. To overcome this limitation, a ML classifier was trained to predict outcome classes using only the four process parameters. An XGBoost model achieved strong predictive performance, with a test set F1 score exceeding 0.9. The classifier's probabilistic output defines the stability component of the fitness score as the predicted probability of the "good" class. Figure 10 (b) illustrates probability isolines in the melt temperature–casting speed plane, highlighting regions where stability exceeds 0.95.



(a) (b)

Figure 10. (**a**) Overview of the proposed ML workflow and key results.; (**b**) Illustration of the fitness score formulation used for ranking generated process parameter suggestions.

The product quality component of the fitness score is defined as the fraction of the predicted grain-size distribution falling within two predefined thresholds, which is to be maximized. In the unimodal configuration, these thresholds and corresponding quality features must be specified prior to model training. By contrast, the bimodal configuration provides greater flexibility, as quality criteria (e.g., imposing an upper bound on maximum grain size) can be modified dynamically through adjustments to the fitness function without retraining the model. In both unimodal and bimodal settings, generating 100 candidate solutions prior to filtering requires approximately 0.5 seconds on average, satisfying real-time operational requirements.

6. Discussions and Challenges

The deployment of reduced-order and physics-informed, data-driven process models within digital advisory systems, digital shadows, and digital twins is hindered by several fundamental and interrelated challenges. Foremost among these are data availability and data quality, as many manufacturing processes lack sufficiently rich, well-distributed, and high-resolution datasets to span their full multidimensional parameter spaces. Sparse, noisy, or inconsistent process data undermine model robustness, predictive accuracy, and generalizability—limitations that become especially critical for transient, multi-physical, or multi-scale processes. Furthermore, achieving adequate and balanced coverage of the parameter space remains difficult. Even when data is available, it is often unevenly distributed across normal, near-boundary, and extreme operating regions, leaving real-time advisory systems vulnerable to "blind spots" and degraded predictive performance when confronted with previously unseen conditions.

A further challenge lies in data-model fitting, as reduced-order and physics-informed models frequently struggle to capture steep gradients, phase transitions, abrupt nonlinear behaviors, and complex initial or boundary conditions. These issues are particularly pronounced in processes characterized by rapid thermal fluctuations, solidification fronts, or large deformation zones. Compounding these difficulties are rate-dependency effects, since many manufacturing operations (including vertical, horizontal, and high-pressure die-casting processes) exhibit rapid heating and cooling cycles, or repetitive cyclic behavior, that often lead to time-shifted, unstable, or inaccurate ROM predictions. Together, these challenges highlight the inherent complexity of integrating high-fidelity data-driven models into advanced process digitalization frameworks.

The integrating ROM-based predictive models into digital process advisory framework adds an additional layer of difficulty. These frameworks require near-instantaneous predictions, high reliability across the entire operational window, and compatibility with heterogeneous data sources including experimental measurements, CFD/FE simulations, and live sensor streams. Ensuring semantic consistency through ontologies, enabling continuous model updating, and maintaining physical validity while generating large synthetic datasets (e.g., via ROM or AC-VAE) are all essential yet technically demanding. Collectively, these challenges including spanning data generation, data fidelity, solver/interpolator performance, multi-scale modeling, dynamic domain representation, and real-time integration, define the core limitations that must be addressed for data-driven and reduced-order models to function robustly within modern process digitalization ecosystems.

7. Concluding Remarks

The findings demonstrate that physics-informed, reduced-order data models can significantly accelerate process design, optimization, and decision making in manufacturing, but only when supported by a rigorous data strategy that treats database construction, validation, and continuous updating as core engineering activities. Beginning with calibrated simulations and limited yet high-quality experimental measurements, small, well-balanced snapshot matrices—generated using Sobol or Latin Hypercube sampling—provide the foundation for reliable eigen-based decomposition methods (e.g., SVD/POD) and informed selection of solver–interpolator combinations. Once validated across normal, near-boundary, and extreme DOE scenarios, these models can be scaled confidently to generate large, physically consistent synthetic datasets, enabling advanced learners such as AC-VAEs and supporting real-time advisory functionalities.

At the same time, the research highlights several non-negotiable challenges that must be addressed before such models can be robustly deployed within digital shadows and digital twins. These include ensuring data availability and quality, achieving balanced coverage of the parameter space, and accurately capturing initial and boundary conditions. Additional complexities arise from the need to model steep thermal and mechanical gradients, phase transitions, rate-dependent transients, and multi-scale, multi-physics interactions, all of which must be organized within semantically consistent, evolvable database architectures.

This studies further illustrate the transformative potential of integrating ROM-based mass synthetic data generation with generative machine-learning frameworks such as AC-VAEs. These hybrid pipelines enable the rapid creation of tens of thousands of physically coherent process scenarios, supporting inverse design, uncertainty quantification, and fast advisory inference at a fraction of the computational cost of full CFD or FE simulations. However, the acceleration gained through these methods comes with its own set of challenges: the models must still contend with steep data gradients, abrupt phase changes, rate-dependent fluctuations, evolving numerical domains, and the inherent mismatch in spatial and temporal resolution between macro-scale thermal fields and micro-scale grain evolution. For this reason, rigorous, multi-stage validation procedures remain essential, combining DOE-based testing, robustness checks under boundary conditions, and consistency assessments across transient cycles. These measures ensure that the predictive framework remains accurate, stable, and trustworthy throughout its deployment in advanced digital manufacturing environments

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org; CFD open-source code for numerical simulations (accessed on 15 Feb. 2026).

Author Contributions: S.M.: methodology, investigation, scripting (CVAE), data curation, validation, writing—original draft preparation, proofreading, editing, visualization. A.M.H.: conceptualization, methodology, writing—original draft preparation, software (data models), validation, writing—review, editing, visualization. S.E.: supervision, writing—original draft preparation, proofreading, editing. M.H.: investigation, data curation, validation, writing—original draft preparation, proofreading. R.G.V.: investigation, CFD simulations, validation, writing—original draft preparation, proofreading, visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financially supported by the Austrian Institute of Technology (AIT) under the UF2025 funding program, and by the Austrian Research Promotion Agency (FFG) through the opt1mus project (FFG No. 899054).

Data Availability Statement: The sharing of raw data required to reproduce the case studies are considered upon request by readers.

Acknowledgments: The authors gratefully acknowledge the technical and financial support provided by the Austrian Federal Ministry for Innovation, Mobility and Infrastructure, the Federal State of Upper Austria, and the Austrian Institute of Technology (AIT). Special thanks are extended to David Blacher and Johannes Kronsteiner for their valuable contributions to this research.

Conflicts of Interest: Authors Sofija Milicic, Amir M. Horr, Manuel Hofbauer, and Rodrigo Gómez Vázquez are employed by the company LKR Light Metals Technologies Ranshofen, Austria**.** Author Stefanie Elgeti is employed by Institute of Lightweight Design and Structural Biomechanics, TU Wien, Chair for Computational Analysis of Technical Systems, RWTH Aachen University, Germany, and DCSE, TU Delft, The Netherlands.

References

1. Ciccone, F., Bacciaglia, A., Ceruti, A. "Optimization with AI in AM: a systematic review," *J. Braz. Soc. Mech. Sci. Eng.*, **2023**, https://doi.org/10.1007/s40430-023-04200-2.
2. Chen, K. et al. "A review of ML in AM: design and process," *Int. J. Adv. Manuf. Technol.*, **2024**, https://doi.org/10.1007/s00170-024-14543-2.
3. Chia, H. Y. et al. "Process parameter optimization of metal AM: review & outlook," *J. Mater. Inf.*, **2022**. https://doi.org/10.20517/jmi.2022.18.
4. Menold, T. et al. "Laser material processing optimization using Bayesian optimization," *Light: Adv Manuf*, **2024**. https://doi.org/10.37188/lam.2024.032.
5. Liu, X., Han, X., Wang, H. et al. A modelling and updating approach of digital twin based on surrogate model to rapidly evaluate product performance. *Int J Adv Manuf Technol* **2023,** 129, https://doi.org/10.1007/s00170-023-12646-w.
6. Tang, Y., Sajadi, P., Rahmani Dehaghani, M. et al. A systematic online update method for reduced-order-model-based digital twin, *J Intell Manuf* **2025**, 36, https://doi.org/10.1007/s10845-024-02524-x.
7. Horr, A.M.; Drexler, H. Real-Time Models for Manufacturing Processes: How to Build Predictive Reduced Models. *Processes* **2025**, *13*, 252. https://doi.org/10.3390/pr13010252.
8. Mallor C., Lani S., Zambrano V., Ghasemi-Tabasi H., Calvo S., Burn A., Integrating 3D printing, simulations and surrogate modelling: A comprehensive study on additive manufacturing focusing on a metal twin-cantilever benchmark, *Adv in Ind and Manuf Eng*, 2025, 10, 100162, https://doi.org/10.1016/j.aime.2025.100162.

doi:10.20944/preprints202604.0566.v1

9. Balu A., Sarkar S., Ganapathysubramanian B., Krishnamurthy A., Physics-aware machine learning surrogates for real-time manufacturing digital twin, *Manuf Letters*, **2022**, 34, https://doi.org/10.1016/j.mfglet.2022.08.013.
10. Leng, J., Zuo, K., Xu, C. et al. Physics-informed machine learning in intelligent manufacturing: a review. *J Intell Manuf* **2025**. https://doi.org/10.1007/s10845-025-02641-1.
11. Pan, L., Li, G., Zhu, T., Liu, D., Wang, Y., and Lu, Y. Physics-Informed Machine Learning in Design and Manufacturing: Status and Challenges. *ASME. J. Comput. Inf. Sci. Eng*. **2025,** https://doi.org/https://doi.org/10.1115/1.4070100.
12. Uhrich, B., Pfeifer, N., Schäfer, M. et al. Physics-informed deep learning to quantify anomalies for real-time fault mitigation in 3D printing. *Appl Intell* **2024**, 54, https://doi.org/10.1007/s10489-024-05402-4.
13. Yang S, Peng S, Guo J, Wang F. A review on physics-informed machine learning for monitoring metal additive manufacturing process. *Adv. Manuf*. **2024**, 2, https://doi.org/10.55092/am20240008.
14. Anonyuo, S., Kwakye, J.M., & Ozowe, W., A Review of Quality Control and Process Optimization in High-Volume Semiconductor Manufacturing, World Journal of Engineering and Technology Research, **2024**, 3(2), 22–27, https://doi.org/10.53346/wjetr.2024.3.2.0060.
15. Lu, Z., Ren, N., Xu, X. et al. Real-time prediction and adaptive adjustment of continuous casting based on deep learning. Commun Eng **2023**, 2, 34, https://doi.org/10.1038/s44172-023-00084-1.
16. Knoth, S. (2002). Statistical Process Control. In: Applied Quantitative Finance. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-05021-7_11.
17. Montgomery, D. C. The use of statistical process control and design of experiments in product and process improvement. IIE Transactions, 1992, 24(5), 4–17. https://doi.org/10.1080/07408179208964241.
18. Haridy, S., Gouda, S.A. & Wu, Z. An integrated framework of statistical process control and design of experiments for op-timizing wire electrochemical turning process. Int J Adv Manuf Technol, 2011, 53, 191–207, https://doi.org/10.1007/s00170-010-2828-7.
19. Dogan, O.; Areta Hiziroglu, O. Empowering Manufacturing Environments with Process Mining-Based Statistical Process Control. Machines 2024, 12, 411. https://doi.org/10.3390/machines12060411.
20. Montgomery, D. C. The 100th anniversary of the control chart. Journal of Quality Technology, **2024**, 56(1), 2–4. https://doi.org/10.1080/00224065.2023.2282926.
21. Best M., Neuhauser D., Walter A Shewhart, 1924, and the Hawthorne factory, BMJ Quality & Safety, **2006**, 15:142-143, https://doi.org/10.1136/qshc.2006.018093.
22. Bradford, P.G., & Miranti, P.J., Information in an Industrial Culture: Walter A. Shewhart and the Evolution of the Control Chart, 1917–1954. Information & Culture, **2019**, 54(2), 179-219, https://doi.org/10.7560/IC54203.
23. Luceño, A., *Control Charts. In: Lovric, M. (eds) International Encyclopedia of Statistical Science*. Springer, 2025, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-69359-9_132.
24. Swamidass, P.M., *SHEWHARTCONTROL CHARTS. In: Swamidass, P.M. (eds) Encyclopedia of Production and Manufacturing Management*, 2000, Springer, New York, NY. https://doi.org/10.1007/1-4020-0612-8_874.
25. Radanliev, P., Artificial intelligence: reflecting on the past and looking towards the next paradigm shift. Journal of Experimental & Theoretical Artificial Intelligence, **2025**, 37(7), 1045–1062. https://doi.org/10.1080/0952813X.2024.2323042.
26. Koren Y., Heisel U., Jovane F., Moriwaki T., Pritschow G., Ulsoy G., Van Brussel H., Reconfigurable Manufacturing Systems, CIRP Annals, 1999, 48, 2, 527-540, https://doi.org/10.1016/S0007-8506(07)63232-6.
27. Coons S. A., *An outline of the requirements for a computer-aided design system*, In Proceedings of the May 21-23, 1963, spring joint computer conference (AFIPS '63 (Spring)). Association for Computing Machinery, New York, NY, USA, 299–304. https://doi.org/10.1145/1461551.1461588.
28. Marr D., Hildreth E., *Theory of edge detection*, Proc. Biol. Sci. 1st February **1980**; 207 (1167), 187–217. https://doi.org/10.1098/rspb.1980.0020.
29. Buchanan B. G., Shortliffe E. H., *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley series in artificial intelligence, June 1984, Addison-Wesley Longman Publishing, MA, United States.

30. Zadeh L.A., The concept of a linguistic variable and its application to approximate reasoning-I, Information Sciences, **1975**, 8, 3, 199-249, https://doi.org/10.1016/0020-0255(75)90036-5.
31. Kanet J. J., Adelsberger H. H., Expert systems in production scheduling, European Journal of Operational Research, 1987, 29, 1, 51-59, https://doi.org/10.1016/0377-2217(87)90192-5.
32. slam, M.T.; Sepanloo, K.; Woo, S.; Woo, S.H.; Son, Y.-J. A Review of the Industry 4.0 to 5.0 Transition: Exploring the Inter-section, Challenges, and Opportunities of Technology and Human–Machine Collaboration. Machines 2025, 13, 267. https://doi.org/10.3390/machines13040267.
33. Oks, S.J., Jalowski, M., Lechner, M. et al. Cyber-Physical Systems in the Context of Industry 4.0: A Review, Categorization and Outlook. *Inf Syst Front* **2024**, 26, https://doi.org/10.1007/s10796-022-10252-x.
34. Katherine van-Lopik, Steven Hayward, Rebecca Grant, Laura McGirr, Paul Goodall, Yan Jin, Mark Price, Andrew A. West, Paul P. Conway, A review of design frameworks for human-cyber-physical systems moving from industry 4 to 5, *IET Cyber-Physical Systems: Theory & Applications* **2023**, https://doi.org/10.1049/cps2.12077.
35. Krugh, M., Mears, L.: A complementary cyber-human systems framework for industry 4.0 cyber-physical systems. *Manuf. Lett*. **2018**, 15, https://doi.org/10.1016/j.mfglet.2018.01.003.
36. Van-Lopik, K. Considerations for the design of next-generation interfaces to support human workers in Industry 4.0, *Doctoral Thesis*, Loughborough University, 2019, https://doi.org/10.26174/thesis.lboro.8132069.v1.
37. Nakagawa, E.Y., et al.: Industry 4.0 reference architectures: state of the art and future trends. *Comput. Ind. Eng*. **2021**, 156, 107241 https://doi.org/10.1016/j.cie.2021.107241.
38. Sharpe, R., et al.: An industrial evaluation of an Industry 4.0 reference architecture demonstrating the need for the inclusion of security and human components. *Comput. Ind*. **2019**, 108, https://doi.org/10.1016/j.compind.2019.02.007.
39. Nguyen, H., Ganix, N., Ion, L.: Human - centred design in industry 4. 0: case study review and opportunities for future research. J. Intell. Manuf. **2022**, 33 https://doi.org/10.1007/s10845-021-01796-x.
40. Manzoni, A., Quarteroni, A. & Rozza, G. Computational Reduction for Parametrized PDEs: Strategies and Applications. Milan J. Math. **2012**, 80, 283–309, https://doi.org/10.1007/s00032-012-0182-y.
41. Dissanayake, M.W.M.G. and Phan-Thien, N., Neural-network-based approximations for solving partial differential equations. Commun. Numer. Meth. Engng., **1994**, 10: 195-201. https://doi.org/10.1002/cnm.1640100303.
42. Lagaris I. E., Likas A., Fotiadis D. I., "Artificial neural networks for solving ordinary and partial differential equations," in IEEE Transactions on Neural Networks, **1998**, 9, 5, 987-1000, https://doi.org/10.1109/72.712178.
43. Psichogios, D.C. and Ungar, L.H., A hybrid neural network-first principles approach to process modeling. AIChE J., **1992**, 38: 1499-1511. https://doi.org/10.1002/aic.690381003.
44. Raissi M., Perdikaris P., Karniadakis G.E., Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Computational Physics, **2019**, 378, 686-707, https://doi.org/10.1016/j.jcp.2018.10.045.
45. Rouhollahi, A., Rismanian, M., Ebrahimi, A., Ilegbusi, O. J., Nezami, F. R., Prediction of directional solidification in freeze casting of biomaterial scaffolds using physics-informed neural networks, Biomedical Physics & Engineering Express, 2024, 10, 6, https://doi.org/10.1088/2057-1976/ad7960.
46. Würth T., Krauß C., Zimmerling C., Kärger L, Physics-informed neural networks for data-free surrogate modelling and engineering optimization – An example from composite manufacturing, Materials & Design, **2023**, 231, 112034, https://doi.org/10.1016/j.matdes.2023.112034.
47. Song R., Eo S., Lee M., Lee J., Physics-informed machine learning for optimizing the coating conditions of blade coating. Physics of Fluids, **2022**; 34 (8): 082112, https://doi.org/10.1063/5.0098666.
48. Horr, A.M. Notes on New Physical & Hybrid Modelling Trends for Material Process Simulations. *J. Phys. Conf. Ser.* **2020**, *1603*, 012008. https://doi.org/10.1088/1742-6596/1603/1/012008.
49. Horr, A.; Blacher, D.; Vázquez, G. On Performance of Data Models and Machine Learning Routines for Simulations of Casting Processes. *BHM Berg-Und Hüttenmännische Monatshefte* **2025**, *170*, 28–36. https://doi.org/10.1007/s00501-024-01537-6.

 doi:10.20944/preprints202604.0566.v1

50. Horr, A.M. Real-Time Modeling for Design and Control of Material Additive Manufacturing Processes. *Metals* **2024**, *14*, 1273. https://doi.org/10.3390/met14111273.
51. Horr, A.M.; Hartmann, M.; Haunreiter, F. Innovative Data Models: Transforming Material Process Design and Optimization. *Metals* **2025**, *15*, 873. https://doi.org/10.3390/met15080873.
52. Horr, A.M. Real-time Modelling and ML Data Training for Digital Twinning of Additive Manufacturing Processes. *BHM Berg-Und Hüttenmännische Monatshefte* **2024**, *169*, 48–56. https://doi.org/10.1007/s00501-023-01416-6.
53. Kingma, D.P.; Welling, M. An Introduction to Variational Autoencoders. Found. Trends Mach. Learn. **2019**, 12, 307–392. https://doi.org/10.1561/2200000056.
54. Doersch, C. Tutorial on Variational Autoencoders. arXiv **2016**, arXiv:1606.05908. Available online: https://arxiv.org/abs/1606.05908 (accessed on 3 March 2026).
55. Ivanov, O.; Figurnov, M.; Vetrov, D. Variational Autoencoder with Arbitrary Conditioning. arXiv **2018**, arXiv:1806.02382. Available online: https://arxiv.org/abs/1806.02382 (accessed on 3 March 2026).
56. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv **2015**, arXiv:1505.04597. Available online: https://arxiv.org/abs/1505.04597 (accessed on 3 March 2026).
57. Bennon W.D., Incropera F.P., A continuum model for momentum, heat and species transport in binary solid-liquid phase change systems—I. Model formulation. Int. J. Heat and Mass Transfer, **1987**, 30, 10, 2161-2170, https://doi.org/10.1016/0017-9310(87)90094-9.
58. Lebon B., directChillFoam: an OpenFOAM application for direct-chill casting. J. Open Source Software, **2023**, 8(82), 4871, https://doi.org/10.21105/joss.04871.
59. Thévoz Ph., Desbiolles J.L, Rappaz M., Modelling of equiaxed microstructure formation in casting. Metallurgical and Materials Transactions A, **1989**, 20, 311–322, https://doi.org/10.1007/BF02670257.