Data-driven modeling of yield surfaces in heterogeneous microstructures using computer science methods*

Valeriia Hritskova[1,*,†], Oleksii Vodka[1,*,†], Mariia Shapovalova[1,*,†] and Oleh Semenenko[4,†]

[1] National Technical University «Kharkiv Polytechnic Institute», 2, Kyrpychova str., Kharkiv, 61002, Ukraine

Abstract

Rapid advances in computational mechanics have necessitated a transition from phenomenological models to precision-based computational frameworks that account for internal structural heterogeneity. This paper presents a data-driven computational approach for identifying and reconstructing the actual yield surfaces of polycrystalline materials. The methodology is based on an integrated pipeline that combines voxel-based microstructure generation via cellular automata in MatViz3D with automated finite element analysis in the ANSYS environment. A specialized computational core is implemented to manage extensive stress-strain state datasets, utilizing hierarchical HDF5 storage and vectorized tensor processing to ensure high performance. Key algorithmic developments include an Eigen-tracking method for principal axis stabilization and a probabilistic analysis of slip system activation based on Schmid's law. The final reconstruction of the yield surface is achieved using Kernel Density Estimation (KDE) to provide a continuous probabilistic representation of the plastic transition boundary. The analysis identified a mean Schmid factor of 277 MPa (ratio 0.92, deviation 0.03) and captured faceted geometries with specific modes for uniaxial (312.9–314.7 MPa), shear (179.5 MPa), and equi-biaxial loading (448.2 MPa). The proposed framework enables the quantitative assessment of stochastic anisotropy and discrete plastic flow effects without a priori specification of the yield function's form. To support the reproducibility and further development of data-driven models, the generated numerical datasets are made publicly available via the Zenodo repository.

Keywords

Data-driven modeling, Polycrystalline materials, Yield surface reconstruction, Kernel density estimation, Finite element analysis, Cellular automata, High-performance data processing

1. Introduction

Rapid advances in computational technology and the availability of large datasets have established materials informatics as a key field for accelerating material discovery [1, 2]. Data-driven approaches effectively identify complex relationships between microstructure and macroscopic properties that traditional analytical models fail to capture [3, 4], emphasizing the importance of FAIR data principles [5]. Machine learning (ML) has become central to these workflows, utilized for high-dimensional design optimization [6], surrogate modeling [7], and the extraction of microstructural descriptors using vision transformers [8, 9].

Modern development in computational materials modeling is characterized by a transition toward integrated intelligent systems covering the full cycle from image segmentation to predicting mechanical properties. Recent advances in deep learning have enabled the implementation of transformer-based architectures, such as MicroStructFormer [10], for the comprehensive description of heterogeneous environments. Particular attention is warranted by the evolution of generative models: from variational autoencoders (VAE) [11] and StyleGAN2 [12]

* CMIS-2026: Eighth International Workshop on Computer Modeling and Intelligent Systems, May 5, 2026, Zaporizhzhia, Ukraine

[1,*] Corresponding author.

[†] These authors contributed equally.

valeriia.hriskova@infiz.khpi.edu.ua (V. Hritskova); oleksii.vodka@khpi.edu.ua (O. Vodka); mariia.shapovalova@khpi.edu.ua (M. Shapovalova); oleh.semenenko@infiz.khpi.edu.ua (O. Semenenko)

0009-0002-6825-726X (V. Hritskova); 0000-0002-4462-9869 (O. Vodka); 0000-0002-4771-7485 (M. Shapovalova); 0009-0003-7275-2412 (O. Semenenko)

networks to modern diffusion models [13, 14], which provide an unprecedented level of topological control in the synthesis of three-dimensional microstructures.

In computational mechanics, this paradigm replaces classical constitutive models with the direct use of simulation data [15], employing deep neural architectures to predict stress fields and mechanical behavior in complex polycrystalline materials [16, 17]. A fundamental concept in constitutive modeling is the yield surface, which defines the elastic-plastic boundary . Classical criteria like von Mises or Tresca often fail for materials with complex microstructures due to assumptions of isotropy [18, 19]. Consequently, physics-informed hybrid approaches [20, 21] and specifically physics-informed neural networks (PINNs) [22, 23, 24] are increasingly applied to reproduce complex anisotropic yield surfaces. This integration of mechanical constraints directly into the learning process is critical for multiscale modeling [25, 26], defining a transition from purely empirical models to physically substantiated data-centric frameworks [23, 27].

Within the framework of Integrated Computational Materials Engineering (ICME), standardized digital microstructures are essential for interoperability [28]. Various algorithmic approaches, including cellular automata-based frameworks, generate synthetic polycrystalline datasets [29, 30, 31, 32]. Sensitivity studies indicate that nucleation and growth parameters significantly influence resulting statistical characteristics [33], while image processing and statistical methods facilitate the automated extraction of morphological descriptors [34, 35, 36].

Despite substantial progress, most existing studies focus on predicting individual material properties or local stress fields. The problem of reconstructing the complete yield surface as a geometric object in stress space remains less explored. Moreover, many current approaches do not explicitly account for the statistical nature of plastic deformation mechanisms, including the distribution of crystallographic orientations and associated Schmid factors.

The present work proposes a computational approach for analyzing the formation of yield surfaces in polycrystalline materials based on digital microstructure generation, statistical analysis of grain orientations, and mechanical response modeling. This approach enables the investigation of the influence of microstructural characteristics on the geometry of yield surfaces and provides a foundation for further integration of data-driven methods into computational materials mechanics.

2. Problem Statement

The modern paradigm of computational mechanics and the modeling of physical media necessitates a transition from generalized phenomenological models to precision-based computational frameworks that account for internal structural heterogeneity. A critical challenge in modeling the nonlinear behavior of materials arises from the fact that classical analytical criteria (e.g., von Mises) rely on the assumption of material homogeneity. This assumption leads to significant numerical inaccuracies when calculating local stress fields in complex polycrystalline topologies and results in the erroneous prediction of the plastic transition point due to the neglect of the stochastic nature of slip system activation.

The objective of this work is to develop an algorithmic framework and implement a software pipeline for the identification of the actual yield surface of heterogeneous systems. This approach is based on the determination of the probabilistic characteristics of Schmid factors and the multi-scale analysis of the mechanical response of structural elements.

To achieve this objective, the following tasks must be addressed:

1. Develop an automated computational workflow for the programmatic translation of voxel-based object geometry (generated via cellular automata in MatViz3D) into the ANSYS finite element analysis (FEA) environment, with subsequent data consolidation into structured HDF5 binary arrays.
2. Formulate a methodology for identifying the probabilistic characteristics of Schmid factors for an ensemble of structural units. This involves creating a tensor processing core to

transform stress fields from the global coordinate system to local orientation systems and computing the statistical distribution of resolved shear stresses.

3. Perform algorithmic identification of boundary surfaces based on the generated numerical datasets. This task entails reconstructing the yield surface in the principal stress space as a complex polyhedron representing the envelope of activated slip systems, incorporating the identified probabilistic patterns of orientation distribution.
4. Perform statistical validation of the identification results by correlating the numerical distributions with classical analytical criteria (e.g., von Mises limit) to evaluate the framework's stability and ensure the physical consistency of the reconstructed yield surfaces.

The scientific novelty of the results lies in the development of a data-driven identification approach that moves beyond the traditional a priori definition of the yield function. By determining the probabilistic characteristics of Schmid factors, this framework enables the quantitative assessment of stochastic anisotropy and discrete plastic flow effects within the mesostructure, significantly enhancing the precision of predictive models for heterogeneous media.

3. Software Framework and Algorithmic Implementation

The implementation of the proposed approach is based on the creation of an end-to-end computational pipeline that combines methods of algorithmic microstructure generation, automated finite element analysis, and advanced data management technologies for large-scale datasets.

3.1. Algorithmic Microstructure Generation in MatViz3D

To model the representative microstructures of heterogeneous materials, the MatViz3D software suite employs the cellular automata method. In contrast to traditional vector methods, this approach generates voxel geometry directly to simulate physical crystallization (Fig. 1a).

The logic of grain growth is based on the Moore neighborhood analysis. In three-dimensional space, each active cell (seed) interacts with 26 neighboring voxels. According to the algorithmic framework developed based on previous studies [13, 14, 15], the structure formation process includes several key stages: initialization of a regular voxel grid, random nucleation of crystallization centers, and subsequent iterative growth. Parallel processing via multi-threading enhances computational efficiency for large data arrays.

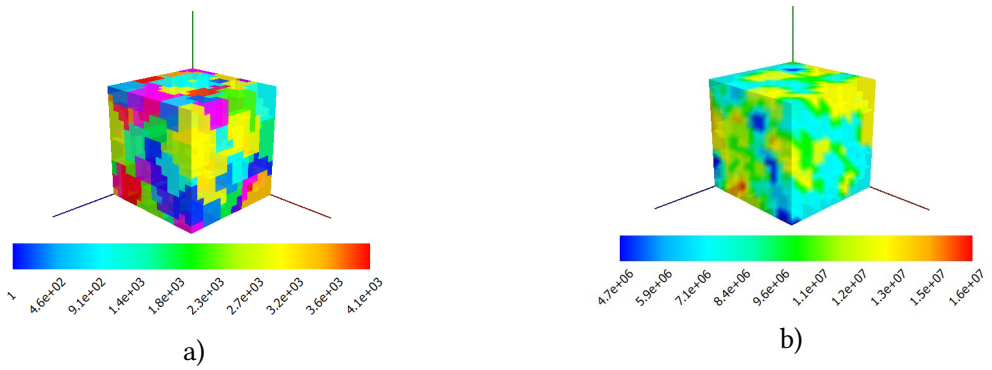

Figure 1: Representative visualization of the computational pipeline results: a) synthetic voxel-based microstructure generated via cellular automata in MatViz3D; b) local equivalent stress distribution within the polycrystalline volume obtained via finite element analysis.

At each step, the algorithm performs a neighborhood analysis for each cell, checking the boundary conditions and the state of neighbors within the Moore radius. If an unfilled cell borders an active grain, it changes its state (ID), simulating the accretion process. After each step, local data from all threads is consolidated into a global array. The iterative space-filling process continues until the computational domain is completely filled, allowing for the reconstruction of complex polycrystalline topologies while accounting for the stochastic distribution of nucleation sites.

3.2. Automation of the Computational Pipeline and Integration with ANSYS

A key step in the framework is the software translation of the generated digital microstructures into the ANSYS finite element analysis environment (Fig. 1b). Each voxel of the structure is automatically converted into an 8-node hexahedral element (SOLID185/186 type). This ensures precise reproduction of grain boundaries without loss of information during discretization.

Integration features include:

- Automatic property assignment involves assigning each grain a unique spatial orientation, which determines its anisotropic stiffness matrix.
- Load parameterization is implemented via an algorithm that generates trajectories in the principal strain space to obtain a complete picture of the yield surface. The use of the Lode angle allows for uniform coverage of the entire spectrum of stress states (from uniaxial tension to pure shear) around the hydrostatic axis.

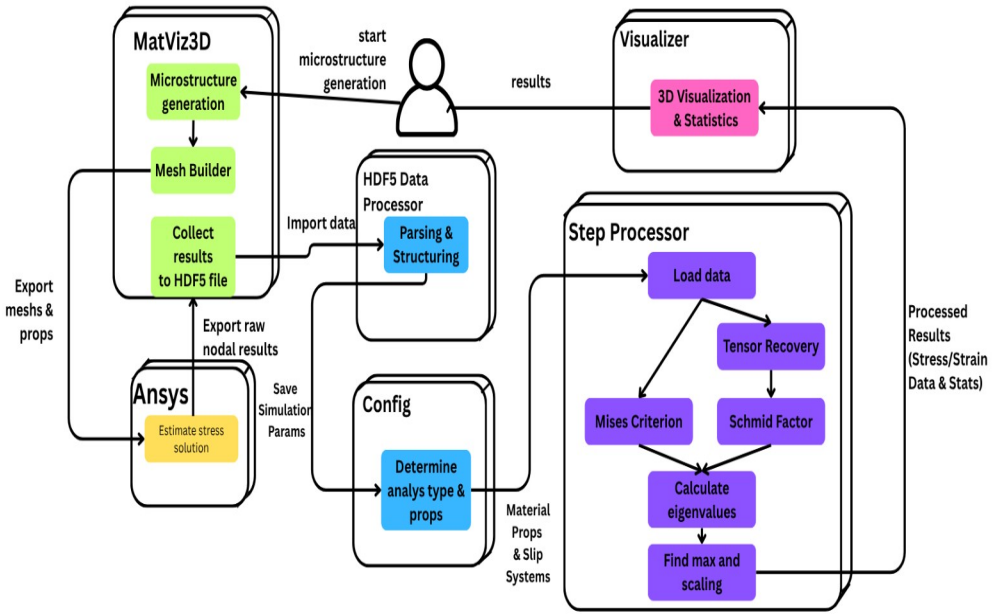

Figure 2: Diagram of software module deployment and data flow.

To detail the functional capabilities of the developed software, a use case diagram was designed to define the system boundaries and interaction scenarios (Fig. 2). The framework provides comprehensive control over the computational process, enabling the configuration of physical constants and visualization parameters, selecting a yield surface identification algorithm, and configuring visualization parameters. The module's final functionality allows researchers not only to perform large-scale stress-strain state calculations but also to conduct interactive statistical analysis of the results, automatically generate graphical reports, and export data for further use in neural networks.

The substantial volume of nodal results generated by these finite element simulations necessitates a specialized data management strategy to ensure computational efficiency and data integrity.

3.3. Hierarchical Structuring of Big Data in HDF5 Format

To efficiently handle extensive volumes of results (large-scale tensor field datasets), a data storage strategy has been implemented in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable). The developed algorithmic framework and the resulting numerical datasets are publicly available in the Zenodo repository [26]. The HDF5 format facilitates hierarchical grouping of geometry, boundary conditions, and loading results within a single binary container, boundary conditions, and load step results within a single binary container (Fig. 3).

To optimize further training of machine learning (ML) models, the data is stored as a "flattened array"—a two-dimensional matrix of dimension $N \times 19$. This structure allows the NumPy and Pandas libraries to perform vectorized reading and processing of tensors without additional transformations, which is critically important for high-volume data analytics in materials science.

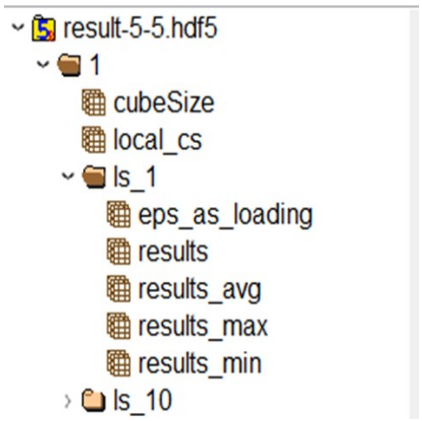

Figure 3: Example of the structure of source HDF5 files.

4. Algorithms for Tensor Processing in the Computing Core

The central component of the developed software is the computational core, which is responsible for transforming raw finite element modeling (FEM) data into physically based plasticity characteristics. The core is based on a combination of methods for efficient large-scale data parsing and algorithms from the mechanics of continuous media.

4.1. Vectorized Processing of Tensor Fields

Processing micromechanical simulation results requires managing extensive stress-strain state (SSS) datasets across numerous loading steps. To ensure high performance, the StepProcessor module implements a strategy of full vectorization of computations based on the NumPy and SciPy libraries.

The main technical solutions and principles of the vectorized approach include:

- the "Flattened Array Strategy";
- elimination of *for* loops;
- memory optimization and Lazy Loading;
- efficient linear algebra.

Implementing these approaches enables high-speed processing of extensive datasets. According to the "Flattened Array Strategy," data from HDF5 is read as a two-dimensional matrix of size, where N is the number of grid nodes. Each row contains a complete data set: a unique identifier, coordinates (x, y, z), a displacement vector (u_x, u_y, u_z), the 6 components of the Cauchy stress tensor, and the 6 components of the elastic strain tensor. This structure optimizes data retrieval for machine learning libraries that process tabular formats efficiently.

Performance is optimized by replacing iterative loops with vectorized array operations. This allows for the use of optimized SIMD instructions and C-compilation within NumPy.

Given the requirements for processing high-volume datasets, a “lazy loading” mechanism has been implemented to optimize memory usage. The software core accesses specific groups and datasets in the HDF5 file only when processing a specific loading step is necessary, which prevents RAM overflow.

Additionally, the np.linalg.eigh method is used to calculate principal stresses and eigenvectors, ensuring efficient linear algebra. Since the stress tensor is symmetric, using a specialized solver instead of general-purpose methods for finding eigenvalues reduces computation time by 30–40% during mass processing of tensor fields.

4.2. Algorithm for Stabilizing Principal Axes (Eigen-tracking)

When analyzing the evolution of the stress-strain state along a loading trajectory, a critical problem arises regarding the numerical instability of the eigenvectors of the stress tensor. Eigenvalues $(\sigma_1, \sigma_2, \sigma_3)$ and their corresponding eigenvectors (principal axes) are determined by linear algebra solvers based on purely mathematical criteria, typically by ascending or descending order of values. However, in materials mechanics, the physical meaning of an axis must be preserved regardless of how the stress magnitude changes along it.

The main causes of destabilization are:

- Crossing of values in cases where, during loading, the values σ_1 and σ_2 become close or swap places; sorting algorithms automatically change the indices of the corresponding vectors, causing the axes to “jump”.
- Sign uncertainty due to the mathematical equivalence of the eigenvectors v and $-v$, as a result of which numerical fluctuations can lead to a sudden inversion of the axis direction to 180°.
- Numerical noise in the form of minor rounding errors in finite element calculations, causing chaotic changes in the orientation of principal axes in regions with low stress gradients.

The computational core utilizes an Eigen-tracking algorithm to minimize geometric divergence between successive loading steps. The logic of the *process_eigen_tracking* algorithm includes the following steps:

1. Calculation of the correlation matrix for each node at the current step t by computing the scalar products C_{ij} between the new eigenvectors $v_i^{(t)}$ and the vectors from the previous step $v_j^{(t-1)}$. A value $C_{ij} \approx 1$ indicates that the vectors describe the same physical axis.
2. Algorithmic reordering based on matrix C analysis to find the optimal permutation that maximizes the diagonal elements. This ensures that the axis that was first at step remains first at step $t-1$, even if its eigenvalue is no longer the largest.
3. Sign alignment after determining the correspondence of vectors, which includes checking the sign and inverting the direction of the new vector if the dot product is negative.
4. Approximation for the initial step with orientation fixed relative to the global coordinate system to create a basis for comparison.

The application of this algorithm allows for obtaining continuous and smooth trajectories in the principal stress space. This is critically important not only for correct 3D visualization of the flow surface, but also for the subsequent use of data in data-driven models, where the gradient stability of the input features directly affects the accuracy of neural network predictions.

4.3. Computational Implementation of Schmid's Law

4.3.1. Theoretical Foundations of Schmid's Law

At the microstructural level, plastic deformation in polycrystals occurs through slip along well-defined crystallographic planes and directions that form slip systems. Each such system α is defined by a pair of unit vectors: the normal to the slip plane n^{α} and the slip direction m^{α} (Fig. 4). According to Schmid's law, slip is activated when the projection of macroscopic stresses onto this system—the Resolved Shear Stress (RSS) τ^{α} - reaches a critical value τ_{crss}.

Mathematically, RSS is calculated as the double convolution of the stress tensor σ and the Schmid orientation tensor P^{α}. First, the orientation tensor of the system is determined:

$$P_{ij}^{\alpha}=m_i^{\alpha}\cdot n_j^{\alpha}, \tag{1}$$

where P_{ij}^{α} denotes the components of the Schmid tensor for the system α, and m_i^{α} and n_j^{α} are the components of the unit vectors of the shear direction and the normal to the slip plane, respectively. Based on this, the desired RSS value is calculated:

$$\tau^{\alpha}=\sigma:P^{\alpha}=\sum_{i=1}^{3}\sum_{j=1}^{3}\sigma_{ij}P_{ij}^{\alpha}, \tag{2}$$

in this equation, τ^{α} represents the reduced tangential stress for a specific system, σ_{ij} represents the components of the stress tensor in the local crystal axes, and the symbol ":" indicates a double tensor contraction.

The actual yield surface of the crystal in this formulation is not smooth (as in the case of the von Mises criterion), but is formed as the envelope of the planes that satisfy the activation conditions for all possible slip systems.

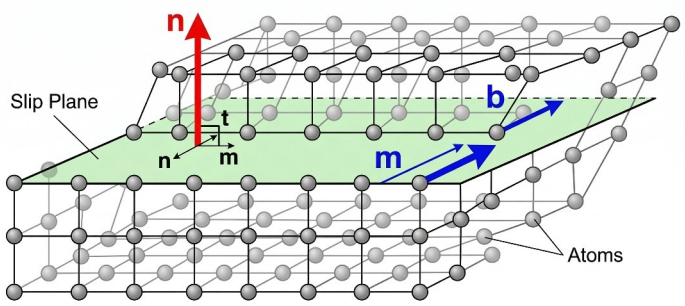

Figure 4: Geometric diagram of a slip system: a plane with normal n and slip direction m in a crystal lattice.

4.3.2. Computational Implementation of Slip System Activation

The software implementation of the *schmid_process_step* method in the data processing module is based on a step-by-step transformation of tensor fields (Fig. 5), taking into account the unique orientation of each grain. The algorithmic pipeline includes the following steps:

1. Reconstruction of spatial orientations based on reading Euler angles $(\varphi_1,\Phi,\varphi_2)$ for each structural unit from the *local_cs* array and constructing a rotation matrix R. This matrix describes the transformation to the local coordinate system of the crystal lattice.
2. Tensor rotation of the stress tensor obtained from the FEM simulation to transform its values into the local grain axes.

3. The algorithm identifies the dominant slip system by iteratively calculating the maximum resolved shear stress τ^{α} across all 12 systems.
4. Data-driven normalization using the calculation of the scaling factor k, which allows scaling the components of the stress-strain matrix and obtaining a point on the actual yield surface without a priori specification of the yield function's form.

This approach allows for the quantitative consideration of the effects of crystallographic anisotropy and the discreteness of plastic flows, providing a foundation for constructing precise faceted flow surfaces, which will be analyzed in detail in the results section.

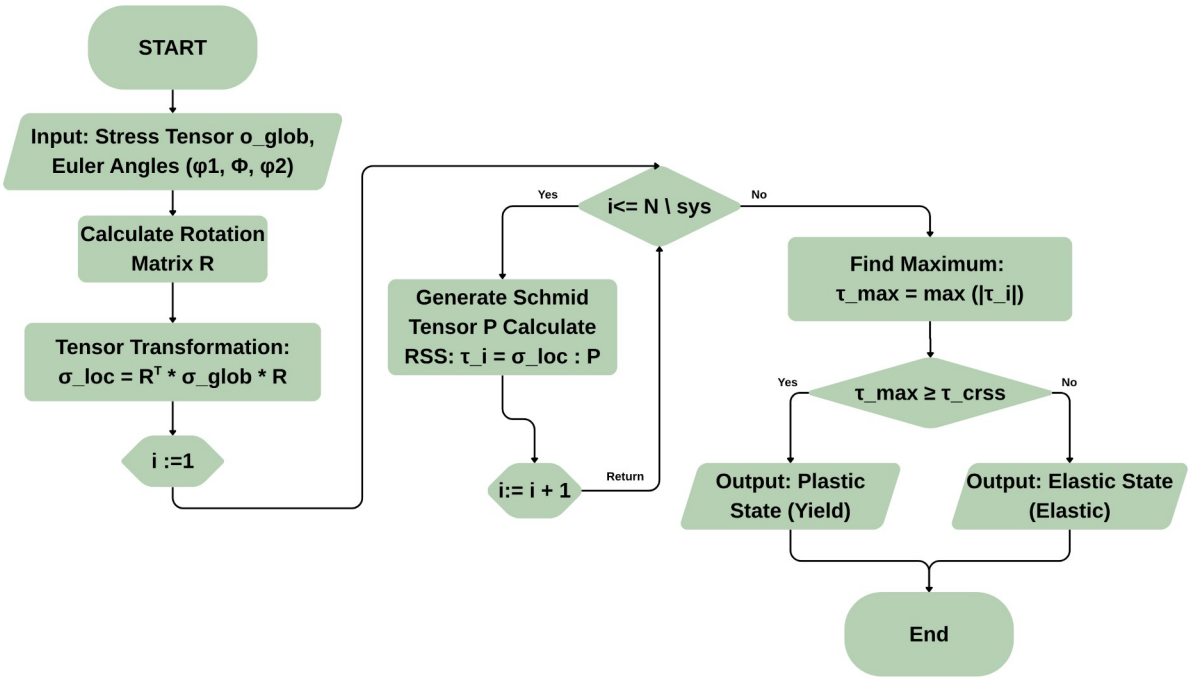

Figure 5: Algorithmic flowchart for calculating the reduced tangential stress and identifying the active slip system.

4.4. Statistical Identification of Surfaces Using the Kernel Density Estimation Method

The Kernel Density Estimation (KDE) method is applied to transform a discrete cloud of deformed state points obtained from FEA simulations into a continuous probabilistic representation of the yield surface. This approach enables the identification of the plastic transition boundary as the locus of maximum local probability density in stress space.

In the proposed framework, the stress state at each point is represented as a six-dimensional vector $\sigma \in \mathbb{R}^6$:

$$\sigma = [\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{yz}, \sigma_{xz}]^T, \tag{3}$$

Directional vectors d_i are utilized to analyze the surface along specific loading trajectories. For instance, uniaxial loading along the X-axis is represented as [1, 0, 0, 0, 0, 0], pure shear XY is represented as [0, 0, 0, 1, 0, 0], and complex states, such as biaxial tension with shear, are represented as [1, 1, 0, 1, 0, 0].

The density estimate $\hat{f}$ for a six-dimensional stress vector σ is defined by the following equation:

$$\hat{f}(\sigma) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\sigma - \sigma_i}{h}\right), \tag{4}$$

where n is the sample size, h is the bandwidth parameter controlling the degree of smoothing, and K is the kernel function (typically Gaussian).

The selection of the bandwidth h is critical for reconstruction accuracy. Instead of utilizing fixed anisotropy coefficients, this study employs Silverman's rule for optimal smoothing in multidimensional space:

$$h=\left(\frac{4}{n(d+2)}\right)^{\frac{1}{d+4}}\hat{s}, \tag{5}$$

Where n denotes the number of points in the sample, $d=6$ is the dimensionality of the stress space, and $\hat{s}$ represents the standard deviation of the stress components. This approach ensures the automatic adaptation of the algorithm to the density of the generated dataset.

Detailed analysis of the yield limit along a selected trajectory d involves computing the marginal probability density. The probability that the yield surface is located at a distance r from the origin in direction d is determined by direct density estimation along the ray.

$$p(r\vee d)\propto\hat{f}_w(r\cdot\alpha_d)\cdot r^5, \tag{6}$$

In this expression, $p(r\vee d)$ represents a one-dimensional probability density function along the ray, and d denotes the direction vector, and α_d denotes the unit direction vector in the six-dimensional stress space. This function yields a distinct peak (mode) corresponding to the physical yield limit for a given complex stress state. This method enables the visualization of yield strength variations and the identification of material anisotropy without the requirement for an a priori analytical yield function.

KDE facilitates automated visualization and analysis of simulation data through two distinct formats. The first format includes a dataset overview, which displays histograms of the distribution of stress components and Mises stress. The second format employs ray marginals to visualize yield strength variations across different stress states and identify the material's physical yield limit.

5. Results

The construction of a reliable model utilizes a representative dataset encompassing the full stress tensor, including principal components and shear fractions. The distribution of input data allows the algorithm to cover the loading space uniformly, which is critical for the precise identification of the yield surface. To ensure physical consistency, the analysis integrates classical criteria: the von Mises energy approach and the calculation of Schmid factors for slip system activation.

Analysis of the simulation results indicates that the calculated stress state points cluster densely near the theoretical yield boundary. This is confirmed by marginal distribution histograms, where primary concentration peaks are localized within the plastic transition zone (Fig. 6). Quantitative indicators of the statistical analysis are summarized in Table 1.

The ratio $\frac{\sigma_{schmid}}{\sigma_y}$ (Ratio) holds particular significance, with a mean value of 0.92. This parameter reflects the correlation between the slip system activation stress identified by the model and the theoretical yield strength of the material. A low standard deviation (0.03) demonstrates high result stability and minimal data dispersion relative to the target surface. Since most values concentrate within the 0.87–1.07 range, the model achieves successful identification of the yield limit while effectively minimizing stochastic errors.

Visual analysis of the dataset confirms these numerical findings. The von Mises distribution comparison (Fig. 6) shows that real stress points (orange) peak sharply near the 270–300 MPa range, aligning with the theoretical limit $\sigma_y=300$ MPa. Projections in the normal stress planes (σ_{xx} vs σ_{yy}) and shear planes (σ_{xy} vs σ_{xz}) visualize comprehensive coverage of the loading space.

The probabilistic characteristics of the yield surface are further detailed via one-dimensional Kernel Density Estimation (KDE) plots (Ray-Marginal), illustrating the probability density $p(r \vee d)$ along specific loading trajectories (Fig. 7).

For uniaxial loading along the X and Y axes, the modes are identified at 312.9 MPa and 314.7 MPa, respectively. In contrast, the Pure Shear XY state exhibits a mode at 179.5 MPa, reflecting the lower stress level required for slip activation under shear conditions. Complex scenarios, such as biaxial loading with shear (mode 250.7 MPa) and equi-biaxial loading (mode 448.2 MPa), demonstrate the model's capacity to capture complex faceted yield surface geometry. The correspondence between the numerical metrics in Table 1 and the visual probability profiles in Fig. 7 confirms the ability of the proposed approach to describe the mechanical behavior of heterogeneous structures while maintaining adherence to physical plasticity criteria.

Table 1
Summary of statistical indicators for yield surface identification

Parameter	Mean	Minimum	Maximum
Schmid factor σ_{schmid} (MPa)	277	260	322
Ratio $\frac{\sigma_{schmid}}{\sigma_y}$	0.92	0.87	1.07
Principal stress S_1 (MPa)	187	−199	671
Normal stress S_x (MPa)	158	-88	689

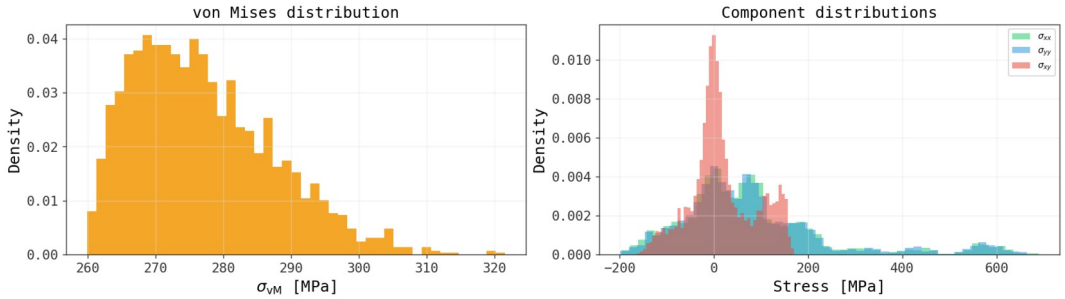

Figure 6: Statistical distribution of the stress-state dataset. Left: Comparison of the calculated von Mises stress distribution with the theoretical yield limit. Right: Density distribution of individual stress tensor components ($\sigma_{xx}, \sigma_{yy}, \sigma_{xy}$) showing the coverage of the computational stress space.

6. Conclusions

The present work implemented an end-to-end computational pipeline that integrates voxel-based microstructure generation in MatViz3D with automated finite element analysis in the ANSYS environment. The developed workflow ensures the seamless translation of representative digital microstructures into structured HDF5 binary arrays, facilitating the management of extensive stress-strain state datasets according to FAIR principles.

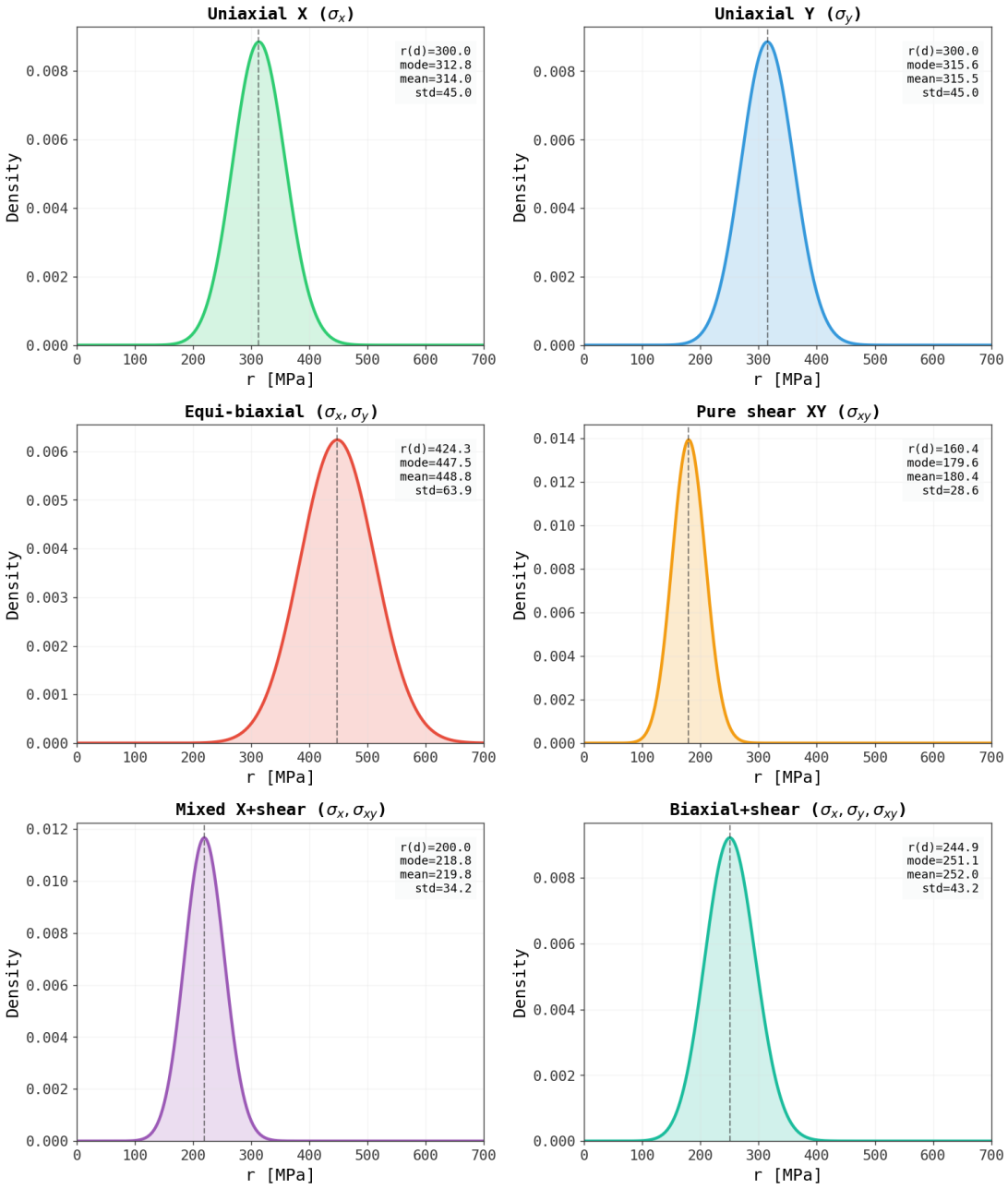

Figure 7: Probabilistic identification of yield limits via 1D Ray-Marginal KDE analysis. Plots show the probability density $p(r \vee d)$ across various loading trajectories (uniaxial, biaxial, and shear states). Each sub-plot indicates the identified mode, mean value, and standard deviation for the respective stress state.

The methodology for identifying the probabilistic characteristics of slip system activation was successfully formulated through a vectorized tensor processing core. Statistical analysis of the mechanical response revealed a mean Schmid factor σ_{schmid} of 277 MPa. The identification of the yield surface was further validated by the $\frac{\sigma_{schmid}}{\sigma_y}$ ratio, which maintained a mean value of 0.92 with a low standard deviation of 0.03. These metrics indicate high stability in the identification process and minimal data dispersion relative to the theoretical plastic transition boundary of 300 MPa.

Algorithmic reconstruction of the boundary surfaces using Kernel Density Estimation (KDE) allowed for the continuous probabilistic representation of the yield surface along diverse loading trajectories. Numerical results identified specific modes for uniaxial loading at 312.9–314.7 MPa, pure shear at 179.5 MPa, and equi-biaxial loading at 448.2 MPa, demonstrating the framework's

ability to capture complex faceted geometries and discrete plastic flow effects. The observed increase of the equi-biaxial yield limit compared to the isotropic 300 MPa baseline reflects the impact of crystallographic constraints and orientation distribution on the macroscopic response.

Despite the robustness of the framework, the current implementation is primarily optimized for specific crystallographic symmetries and does not yet account for grain boundary hardening or thermal effects. Future development will focus on addressing these limitations by incorporating multi-physics constraints and leveraging the advanced generative architectures and transformers discussed in this work to achieve higher predictive fidelity across diverse material classes

This data-driven approach enables the quantitative assessment of stochastic anisotropy in heterogeneous media without the requirement for an a priori specification of the yield function's form. The close correspondence between the identified numerical metrics and visual probability profiles confirms the model's adherence to physical plasticity criteria. This work serves as a foundation for further research into advanced material behavior, the continued expansion of the publicly available materials dataset on Zenodo, and the refinement of regression-based predictive models.

Acknowledgements

This work has been supported by the Ministry of Education and Science of Ukraine in the framework of the realization of the research project “Algorithms, models, and tools of artificial intelligence for two-level modeling of complex materials behavior for dual-use technologies” (State Reg. Num. 0124U000450).

Declaration on Generative AI

During the preparation of this work, the author(s) used Gemini in order to: Grammar and spelling check, and assistance with figure captions. Further, the author(s) used Gemini for Figure 4 in order to: Generate images. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

[1] C. Li, K. Zheng, Methods, progresses, and opportunities of materials informatics, InfoMat 5 (2023). doi:10.1002/inf2.12425.

[2] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, npj Comput. Mater. 3 (2017) 54. doi:10.1038/s41524-017-0056-5.

[3] D. Morgan, R. Jacobs, Opportunities and challenges for machine learning in materials science, Annu. Rev. Mater. Res. 50 (2020) 71–103. doi:10.48550/arXiv.2006.14604.

[4] K.-H. Lee, H. J. Lim, G. J. Yun, A data-driven framework for designing microstructure of multifunctional composites with deep-learned diffusion-based generative models, Eng. Appl. Artif. Intell. 129 (2024) 107590. doi:10.1016/j.engappai.2023.107590.

[5] M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C. Felser et al., FAIR data enabling new horizons for materials research, Nature 604 (2022) 635–642. doi:10.1038/s41586-022-04501-x.

[6] R. Liu, A. Kumar, Z. Chen, A. Agrawal, V. Sundararaghavan, A. Choudhary, A predictive machine learning approach for microstructure optimization and materials design, Sci. Rep. 5 (2015) 11551. doi:10.1038/srep11551.

[7] B. D. Nguyen, P. Potapenko, A. Dermici, K. Govinda, S. Sandfeld, Efficient surrogate models for materials science simulations: machine learning-based prediction of microstructure properties, Mach. Learn. Appl. 16 (2024) 100544. doi:10.1016/j.mlwa.2024.100544.

[8] S. E. Whitman, M. I. Latypov, Machine learning of microstructure–property relationships in materials with vision transformers, Acta Mater. 296 (2025) 121217. doi:10.1016/j.actamat.2025.121217.
[9] M. Shapovalova, O. Vodka, A data-driven approach to the prediction of spheroidal graphite cast iron yield surface probability characteristics, Lect. Notes Netw. Syst. 188 (2021) 565–576. doi:10.1007/978-3-030-66717-7_48.
[10] Zhang W, Chen L, Wang Y. MicroStructFormer: Transformer-based comprehensive microstructure characterization. Materials Horizons, 2025, 12(1): 156-168.
[11] Radev M, Bergmann S, Kopp A. Microstructure generation via variational autoencoders. Computational Materials Science, 2019, 164: 82-92.
[12] Nakamura T, Saito A, Kobayashi H. High-resolution microstructure synthesis using StyleGAN2. npj Computational Materials, 2022, 8(1): 145.
[13] Zhang H, Wang F, Liu J. Stable diffusion for microstructure generation. arXiv preprint arXiv:2305.12345, 2023.
[14] Chen H, Li W, Zhang Y. MicroDiffusion: Topologically controllable microstructure generation. Advanced Materials, 2025, 37(2): 2405678.
[15] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, Comput. Methods Appl. Mech. Eng. 304 (2016) 81–101. doi:10.1016/j.cma.2016.02.001.
[16] Y. Shokrollahi, M. Nikahd, K. Gholami, G. Azamirad, Deep learning techniques for predicting stress fields in composite materials: a superior alternative to finite element analysis, J. Compos. Sci. 7 (2023). doi:10.3390/jcs7080311.
[17] A. L. Frankel, R. E. Jones, C. Alleman, J. A. Templeton, Predicting the mechanical response of oligocrystals with deep learning, Comput. Mater. Sci. 169 (2019) 109099. doi:10.1016/j.commatsci.2019.109099.
[18] N. Anderson, R. Sharan, D. Martin, J. B. Irene, A machine learning model to predict yield surfaces from crystal plasticity simulations, Int. J. Plast. 161 (2023) 103507. doi:10.1016/j.ijplas.2022.103507.
[19] J. N. Fuhg, L. van Wees, M. Obstalecki, P. Shade, N. Bouklas, M. Kasemer, Machine-learning convex and texture-dependent macroscopic yield from crystal plasticity simulations, Materialia 23 (2022) 101446. doi:10.1016/j.mtla.2022.101446.
[20] S. Liu, K. Lee, P. V. Balachandran, Integrating machine learning with mechanistic models for predicting the yield strength of high entropy alloys, J. Appl. Phys. 132 (2022) 105105. doi:10.1063/5.0106124.
[21] J. Peng, Y. Yamamoto, J. A. Hawk, E. Lara-Curzio, D. Shin, Coupling physics in machine learning to predict properties of high-temperature alloys, npj Comput. Mater. 6 (2020) 141. doi:10.1038/s41524-020-00407-2.
[22] Schmidt T, Müller S. Physics-informed neural networks for multiscale modeling of composites. Computational Mechanics, 2024, 73(2): 345-362.
[23] Karniadakis G E, Kevrekidis I G, Lu L, et al. Physics-informed machine learning. Nature Reviews Physics, 2024, 6(1): 22-36.
[24] 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. Journal of Computational Physics, 2019, 378: 686-707.
[25] Brunton S L, Kutz J N. Methods for data-driven multiscale model discovery for materials. Journal of Physics: Materials, 2019, 2(4): 044002.
[26] Jones R, Smith T, Brown A. DeepMaterialNet: Operator learning for multiscale materials modeling. Computer Methods in Applied Mechanics and Engineering, 2022, 398: 115289.
[27] DeCost B, Hattrick-Simpers J, Trautt Z, et al. Data-driven materials science: Status and challenges. JOM, 2023, 75(5): 1423-1436.
[28] G. J. Schmitz, Microstructure modeling in integrated computational materials engineering (ICME) settings: can HDF5 provide the basis for an emerging standard for describing microstructures?, JOM 68 (2016) 77–83. doi:10.1007/s11837-015-1748-2 .

[29] Y. I. Kuchin, R. I. Mukhamediev, K. O. Yakunin, One method of generating synthetic data to assess the upper limit of machine learning algorithms performance, Cogent Eng. 7 (2020) 1718821. doi:10.1080/23311916.2020.1718821 .
[30] M. Abu-Mualla, J. Huang, A dataset generation framework for symmetry-induced mechanical metamaterials, J. Mech. Des. 147 (2024) 1–24. doi:10.1115/1.4066169.
[31] V. Hritskova, O. Semenenko, M. Shapovalova, O. Vodka, Software development for modeling microstructures of polycrystalline materials by cellular automata method, in: Proc. IEEE 5th KhPI Week on Advanced Technology (KhPIWeek), 2024, pp. 1–6. doi:10.1109/KHPIWEEK61434.2024.10877965 .
[32] V. Hritskova, O. Semenenko, M. Shapovalova, O. Vodka, MatViz3D Software Tool for Simulation of Polycrystalline and Hybrid Materials, in: Proc. IEEE 6th KhPI Week on Advanced Technology (KhPIWeek), 2025. doi:10.1109/KhPIWeek61436.2025.11288601.
[33] V. Hritskova, M. Shapovalova, O. Semenenko, O. Vodka, Sensitivity analysis of parameters used to generate material microstructures by cellular automata, Lect. Notes Netw. Syst. 1473 (2025) 118–129. doi:10.1007/978-3-031-94845-9_10 .
[34] V. Hritskova, O. Semenenko, M. Shapovalova, O. Vodka, Algorithms and methods for comparing microstructures of materials based on their images, CEUR Workshop Proc. 3702 (2024) 449–461.
[35] M. Shapovalova, O. Vodka, Image processing technology to determine the parameters of the internal structure of composite materials, in: Proc. IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek), 2021, pp. 539–543. doi:10.1109/KhPIWeek53812.2021.9570099 .
[36] R. Shoghi, A. Hartmaier, Optimal data-generation strategy for machine learning yield functions in anisotropic plasticity, Front. Mater. 9 (2022) 868248. doi:10.3389/fmats.2022.868248.
[37] O. Vodka, O. Semenenko, V. Hritskova, M. Shapovalova, MatViz3D Synthetic 3D Microstructure Dataset of Single Crystal FCC Copper, Zenodo, 2026. doi: 10.5281/zenodo.17865073.
[38] J. Tang, X. Geng, D. Li, Y. Shi, J. Tong, F. Peng, Machine learning-based microstructure prediction during laser sintering of alumina, Sci. Rep. 11 (2021) 10724. doi:10.1038/s41598-021-89816-x.