**DOI: https://doi.org/10.15276/aait.09.2026.11**
**UDC 004.85:539.3**

# Rational data sampling for surrogate modeling of the stress-strain state of plate elements

**Roman L. Onatskyi[1)]**
ORCID: https://orcid.org/0009-0004-6305-8442; roman.onatskyi@infiz.khpi.edu.ua
**Serhii Yu. Misiura[1)]**
ORCID: https://orcid.org/0000-0002-5048-1610; serhii.misiura@khpi.edu.ua. Scopus Author ID: 56416492100
[1)] National Technical University “Kharkiv Polytechnic Institute”, 2, Kyrpychova Str. Kharkiv, 61002, Ukraine

## ABSTRACT

**Background.** Surrogate models based on machine learning are a promising alternative to computationally expensive finite element analysis for predicting the stress-strain state of structural elements. However, their efficiency critically depends on the quality of training data, while most existing approaches rely on abstract stochastic sampling methods that ignore the physical nature of engineering parameters. **Aim.** The aim of the study is to conduct a comparative analysis of the efficiency of stochastic versus physically-oriented data sampling strategies for the surrogate modelling of the stress-strain state of plate elements, with the focus shifted from increasing model complexity to developing an approach for rational dataset formation that integrates a priori engineering knowledge. **Methods.** Using the surrogate modelling of thin rigidly clamped plate deflection as an example, two fundamentally different sampling strategies are compared: the classical strategy, based on Sobol quasi-random sequences, and the proposed rational strategy. The latter is based on the real empirical properties of structural steels and accounts for the standardized plate thicknesses of sheet metal assortments. To evaluate and quantitatively compare the impact of the data structure on the precision of the prediction, an intelligent modelling procedure was developed. The following machine learning algorithms were used as test models for the regression task: K-Nearest Neighbours, Random Forest, XGBoost, Gaussian Process Regression, and Multilayer Perceptron. **Results.** It was found that switching to the rational strategy provides a consistent reduction in mean squared error across all evaluated models by three and a half to four times compared to the baseline strategy. Using the rational dataset increased the coefficient of determination by twenty-seven percent for k-nearest neighbors, twenty-three percent for random forest, and twenty-one percent for XGBoost. The highest precision was achieved by the models of gaussian process regression (with a coefficient of determination of ninety nine percent) and multilayer perceptron (with a coefficient of determination of ninety eight percent). It is proven that the low efficiency of traditional samples is due to the fact that more than fifty percent of the samples contain combinations of geometric dimensions and material properties that do not correspond to any real engineering standards. These unrealistic combinations create “information noise” during training. **Conclusions.** The proposed rational sampling approach serves as a basis for creating robust AI tools for the rapid diagnostics of engineering structures, providing high model generalisation capability with a smaller volume of input data.

**Keywords:** Stress-strain state; surrogate modelling; regression analysis; machine learning; dataset; rational data sampling strategy; rapid diagnostics of engineering structures

***For citation*: Onatskyi R. L., Misiura S. Yu. “Rational data sampling for surrogate modeling of the stress-strain state of plate elements”. *Applied Aspects of Information Technology*. 2026; Vol.9 No.2: 158–171. DOI: https://doi.org/10.15276/aait.09.2026.11**

## INTRODUCTION

Ensuring reliable operation of critical structural elements, such as pipelines, pressure vessels, and storage tanks, is a priority in the mechanical engineering, petrochemical, energy, and aerospace industries. This is particularly relevant within the strategy of transitioning from scheduled repairs to predictive maintenance. This strategy involves forecasting potential failures and assessing the remaining service life of the equipment in real-time, based on the actual condition of the structure, including the presence or evolution of geometric imperfections or operational defects such as corrosion, dents, etc.

A critical requirement for such systems is the accuracy and speed of the calculations. Among traditional methods for determining the stress-strain state (SSS) of complex engineering structures, the finite element method (FEM) is currently the most universal and widespread. Despite its high accuracy and versatility, it is too demanding on computational resources for instant diagnostics and is complex to master.

Surrogate models based on Machine Learning (ML) describe an effective alternative; they are capable of approximating complex nonlinear dependencies of solid mechanics with high speed and can replace complex computations with rapid predictions.

However, the efficiency and reliability of such models are critically dependent on the quality of the training data used in their development. The use of abstract, purely mathematical, synthetic approaches to sample generation often leads to the creation of models unsuitable for real engineering scenarios, as they ignore the physical nature of parameters, treating them as continuous stochastic variables. On the contrary, real engineering tasks operate with standard sheet metal gauges and specific material grades. The practical value of this study lies in the development and verification of a rational approach for forming training datasets that accounts for the physical properties of materials and the design features of objects. In this work, the classical problem of the specific SSS of a square plate is used as a controled benchmark. This is a necessary validation step for the proposed sampling procedure before its application to more complex, resource-intensive tasks, particularly the diagnostics of shell structures with operational defects.

Although, the specific conclusions obtained here can also be used as a basis for creating robust AI tools to diagnose more complex structures, including those with defects. This allows increased accuracy in monitoring systems and the effectiveness of predictive maintenance strategies, ensuring high model generalisation capability with a smaller volume of input data.

Therefore, studies dedicated to the impact of rational data sampling methods on the quality of surrogate modelling of the SSS are current and relevant.

## ANALYSIS OF LITERARY DATA

The integration of artificial intelligence methods into solid mechanics is one of the most dynamic directions of modern research. Traditionally, analytical solutions within the framework of Kirchhoff-Love and Timoshenko-Reissner theories [1], [2] were used to calculate the strength of thin-walled structures, which were subsequently displaced by more universal numerical methods, primarily the finite element method [3], [4]. However, the high computational cost of FEM stimulated the development of surrogate modelling, where, for example, neural networks act as fast approximators of the results of numerical analysis [5], [6], [7].

A critical analysis of scientific publications from recent years allows for the classification of existing approaches to improving the accuracy of such models into three main directions.

A significant portion of modern research focusses on increasing the complexity of neural network architectures. Researchers try to improve prediction accuracy by applying deep convolutional neural networks (CNN), recurrent models, or evolutionary algorithms [8]. For example, in [9], the authors apply genetic algorithms to automatically search for the optimal network topology, and a review study [10] analyses the evolution of architectures from convolutional to generative adversarial networks for structural health monitoring tasks. However, as noted by the authors of [10], excessive complexity of the model often leads to problems with interpretability and robustness. In addition, studies [5], [6] propose complex ensemble methods. However, despite the improved metrics on test samples, such models often demonstrate low robustness when implemented in real engineering systems. This is explained by the fact that within this approach researchers focus primarily on improving the mathematical apparatus of the model, treating training data as a static constant. As a result, optimising architecture on abstract, synthetic, or nonrepresentative data is unable to solve the fundamental problem of model reliability under real operating conditions.

On the other hand, an alternative approach involving physics-informed neural networks (PINNs) [11] is widespread, where differential equations of mechanics are integrated directly into the loss function. This allows models to adhere to physical laws even in the absence of large volumes of data. However, mass implementation of PINNs in engineering practice faces significant barriers related to the difficulty of adapting them to tasks with discrete parameters. Since PINNs operate in a continuous space, accounting for rigid engineering standards, such as fixed sheet thicknesses or specific steel grades, becomes a nontrivial task. Furthermore, formulating loss functions for problems with complex geometry and boundary conditions requires significant computational resources and is often accompanied by convergence problems in gradient methods [12].

The third, less researched but critically important aspect is the quality and representativeness of training data. In most studies, for example, in [13], [14], the formation of the training sample is carried out by generating random (Monte Carlo) or quasirandom (Sobol sequences, Latin Hypercube [15], [16]) parameter combinations. However, as noted by the authors in [17], this approach often fails to adequately cover rare events (distribution tails), which is critical for assessing structural reliability. Developing this direction, the authors of [18] propose using adaptive sampling strategies, where new points are added to the sample iteratively based on model error. This allows for a significant reduction in the volume of necessary data compared to classical random sampling.

Although these approaches represent a step forward, the use of purely mathematical strategies, even adaptive ones, often ignores the physical nature of the research objects. Thus, the use of abstract algorithms inevitably leads to the generation of parameter combinations that lack physical meaning, such as a combination of Young's modulus and Poisson's ratio, which does not correspond to any real material. This creates a paradoxical situation where models expend resources learning physically impossible states, which reduces their sensitivity and accuracy in zones of real engineering solutions.

It is also worth noting that the concept of "data-centric engineering" (DCE) has recently gained increasing weight as an independent paradigm. In particular, in [19], the authors define DCE as the integration of engineering principles with data science to solve complex design and manufacturing problems. They demonstrate that the use of generative models which account for the nature of variability in real processes (e.g., in 3D steel printing) allows for the creation of significantly more accurate digital twins than traditional deterministic approaches. Furthermore, in [20], the authors demonstrate that the use of neural network surrogates within the DCE framework makes global sensitivity analysis practically feasible for complex engineering objects, such as bridges and underground structures, which was previously impossible due to the excessive computational cost of FEM. These works clearly confirm that a focus on the "nature of data", its structure, and correspondence to real physical processes is significant when creating reliable and practically useful surrogate models.

Systematisation of the analysed sources indicates that most researchers focus on complicating model architectures and mathematical apparatus. There is a gap between the discrete, standardised nature of engineering objects and the continuous, stochastic methods of data generation traditionally used for training surrogate models. Currently, there is no approach that allows the integration of a priori engineering knowledge (standard assortments, material properties) directly into the structure of the training sample at the formation stage. Bridging this gap through the development and verification of a rational data sampling strategy is the subject of this study.

## THE PURPOSE AND OBJECTIVES OF THE RESEARCH

The purpose of the study is to conduct a comparative analysis of the efficiency of stochastic versus physically-oriented data sampling strategies for the surrogate modeling of the stress-strain state of plate elements. The study focuses on determining how integrating discrete constraints (standard assortment gauges and real material properties) into the dataset structure affects the predictive accuracy of machine learning models compared to classical quasi-random sequences.

To achieve this goal, the following objectives must be met:

1. Formalise a system of geometric and physical constraints to exclude non-physical states from the parameter space for the specific problem of a square plate subjected to a uniformly distributed load.
2. Develop algorithms and software for the automated generation of training samples using two approaches: a synthetic approach with uniform parameter distribution and a proposed rational approach based on discrete engineering parameters and properties of structural materials.
3. Generate two independent datasets supplemented with the results of FE numerical experiments for deflection at the plate centre.
4. Train and evaluate a complex of surrogate models (KNN, Random Forest, XGBoost, GPR, and MLP) on both datasets to assess prediction stability and error metrics.
5. Perform a comparative analysis of efficiency based on statistical accuracy criteria: mean squared error (MAE), mean absolute error (MAE), coefficient of determination $R^2$, and feature importance analysis.

The practical significance of the work lies in demonstrating a data-centric approach for creating robust AI tools for the diagnostics of engineering structures, which allows for high model generalization capability with a smaller volume of input data.

## RESEARCH METHODOLOGY

The study is based on the hypothesis that integrating a priori knowledge about the discrete nature of engineering parameters into the structure of the training sample will increase the accuracy of surrogate modelling compared to classical stochastic methods.

**Problem statement of surrogate modelling.** The object of study is the substitution of computationally expensive traditional analysis with rapid machine learning predictions. In simplified formulaic form, the calculation of the maximum deflection of a specific square plate under pressure (see Fig. 2) using the FEM is represented as a numerical model $M_{FEA}$. This model implements a mapping from the 5-dimensional input parameter space $X=(a, t, E, \nu, p)$ to the scalar output space $w_{\max}$ (maximum deflection):

 ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

$$w_{max} = M_{FEA}\left(X\right). \quad (1)$$

The $M_{FEA}$ model (1) is sufficiently accurate, but computationally expensive, making its use in optimisation tasks or real-time analysis impossible.

Therefore, the problem is formulated as the development of an alternative surrogate model $f_\theta$, which, using ML, is capable of rapidly and accurately approximating the functional dependence found in FEM (2).

$$f_\theta(X) = w_{\max}. \quad (2)$$

Formally, the ML task consisted of finding the optimal parameters $\theta^*$ of the model $f_\theta$.

That is, having the training dataset

$$D = \{(x_i, w_i)\}_{i=1}^{N}, \quad (3)$$

obtained by parameter generation and FEM calculation, the optimal parameters $\theta^*$ were sought as:

$$\theta^* = \arg\min_\theta \frac{1}{N}\sum_{i=1}^{N}\left(L(f_\theta(x_i), w_i)\right), \quad (4)$$

where $L$ is the loss function.

To quantitatively evaluate the accuracy and efficiency of the formulated prognosis task, three standard statistical metrics are defined for the test datasets: MSE to penalize large outliers, MAE for average absolute deviation in engineering units (mm), and the coefficient of determination $R^2$ to assess overall approximation quality.

Since the input parameters had different physical dimensions and value ranges (e.g., Young's modulus is measured in $10^5$ MPa and higher, while thickness is in millimetres), direct use of initial unnormalized values led to gradient descent instability. To eliminate this problem, a normalisation procedure was applied for each input parameter.

The entire dataset was divided into training and test subsets in an 80 % to 20 % ratio. To ensure statistical reliability, the division into subsets was performed randomly using the scikit-learn library.

**Concept of a rational dataset.** Unlike the widely used standard approach that involves random or quasi-random sequences (Sobol), where all parameters change independently and continuously, the rational feature selection strategy proposed in this work is based on three levels of constraints specified in Table 1.

The proposed list of constraints is not exhaustive; however, for this stage of research, it is accepted as a starting point. The specified in Table 1 criteria cover the main physical-geometric aspects of the problem, which is sufficient for a comparative analysis of predictive model efficiency.

*Table 1.* **Accepted system of constraints for the feature space of the rational dataset**

| Constraint Type | Range | Engineering Meaning |
|---|---|---|
| Material Science | $(E, \nu) \in DB_{materials}$ | Use of real physical-mechanical properties of steels |
| Technological | $t \in \{1.0, 1.2, 1.5, ..., 30.0\}$ | Correspondence to standard sheet metal assortment |
| Geometric | $20 \le a/t \le 100$ | Compliance with the conditions for the applicability of thin plate theory applicability [1] |

***Source*: compiled by the authors**

**Feature space formation and data generation.** For the comparative analysis, two independent datasets of 3000 samples each were formed. The input feature vector for both cases consisted of five parameters: plate width $a$, thickness $t$, Young's modulus $E$, Poisson's ratio $\nu$, and pressure value $p$. The output parameter was the maximum deflection in the centre of the plate $w$.

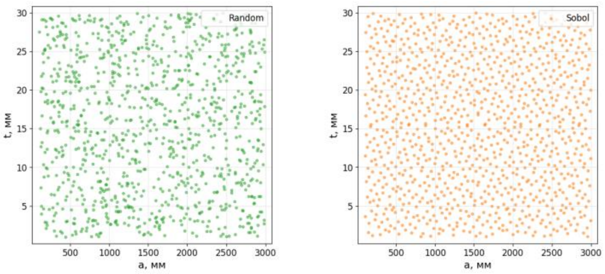
The generation parameters had the following boundaries: width $a \in [100, 3000]$ mm; thickness $t \in [1.0, 30]$ mm; Young's modulus $E \in [10000, 219700]$ MPa; Poisson's ratio $\nu \in [0.2, 0.4]$; load $p \in [1000, 10000]$ Pa.

The actual maximum deflection values $w$ for both datasets were obtained by direct modelling using the FEM analysis in the ANSYS environment.

**Synthetic dataset (DS1). Baseline.** The first dataset used as a baseline for evaluating sampling efficiency is termed the *Synthetic Dataset (DS1)*. This dataset was generated using quasirandom Sobol sequences. This method was chosen to ensure uniform filling of the 5-dimensional parameter hypercube and to avoid point clustering characteristic of pseudo-random algorithms.

It should be noted that initial attempts to generate data using a "naive" pseudorandom approach (based on Monte Carlo type algorithms), using the built-in numpy.random module from the Python environment, led to the appearance of undesirable artefacts, namely clusters and gaps in the parametric space. To overcome this problem and ensure maximal uniform coverage of the space, it was decided to subsequently apply the approach common in engineering modelling based on lowdiscrepancy quasi-random sequences, specifically Sobol sequences. This deterministic method avoids specific biases and obtains a more representative sample.

The difference in value distribution for the pseudorandom algorithm and Sobol sequences can be visually assessed in Fig. 1.




a b

*Fig. 1.* **Comparison of sample spaces for parameter generation methods: a – random method; b – Sobol quasirandom method**

***Source*: compiled by the authors**

In the figure, both panels have identical axis limits (x-axis: plate width a, mm; y-axis: thickness t, mm) for a correct visual comparison. Marker colours correspond to the generation method.

**Rationally formed dataset (DS2).** The second dataset, termed the **physically grounded or rational dataset (DS2)**, was formed taking into account the discrete nature of real engineering material properties and parameters, and in accordance with the criteria specified in Table 1.

According to material science constraints, instead of a continuous distribution of material properties, a fixed set of structural steels and alloys was used (see Table 2). The data source used was the open dataset “Materials and Their Mechanical Properties” [21], which contain material properties from the Autodesk library. In total, the reviewed dataset contains 1552 unique records, 40 % of which correspond to the ANSI standard and 21 % to GOST; the remainders are EN and other standards.

The filtration and unique record selection algorithm relied on the premise that only materials with unique pairs of Young's modulus $E$ and Poisson's ratio $\nu$ were of interest. For this, the working range was limited to $E \in [1.2, 2.197]\cdot 10^5$ MPa and $\nu \in [0.2, 0.4]$, and **the resulting sample was filtered out of duplicates**.

As a result of applying the algorithm, 36 unique representative grades (structural steels, cast irons, alloys) were selected from 1552 records, ensuring the physical connectivity of parameters $E$ and $\nu$. Subsequently, to form the dataset from the materials listed in the table, a sample was chosen randomly and supplemented with parameters according to other criteria.

According to the technological constraint, the thickness values $t$ were uniformly selected from a discrete set in the range $t \in [1.0, 30]$ mm, corresponding to the standard assortment of sheet metal.

*Table 2.* **Unique materials used to form the DS2 dataset**

| | Material | E, МПа | ν |
|---|---|---|---|
| 1 | Steel SAE 1015 | 207000 | 0.3 |
| 2 | Steel SAE 30201 | 193000 | 0.3 |
| 3 | Steel SAE 51403 | 200000 | 0.3 |
| 4 | Grey cast iron class 50 | 124000 | 0.25 |
| 5 | Grey cast iron class 60 | 137000 | 0.25 |
| 6 | Malleable cast iron | 172000 | 0.27 |
| 7 | Nodular cast iron | 159000 | 0.2 |
| 8 | Manganese Bronze C86100 | 120000 | 0.31 |
| 9 | Aluminum Bronze C95200 | 120000 | 0.32 |
| 10 | EN 11SMn30 | 206000 | 0.3 |
| 11 | Nodular cast iron | 169000 | 0.2 |
| 12 | Malleable cast iron | 160000 | 0.27 |
| 13 | Steel ST3KP GOST 380-94 | 201000 | 0.29 |
| 14 | Steel 08 GOST 1050-88 | 203000 | 0.3 |
| 15 | Steel 20 GOST 1050-88 | 212000 | 0.29 |
| 16 | Steel 25 GOST 1050-88 | 198000 | 0.29 |
| 17 | Steel 50 GOST 1050-88 | 216000 | 0.3 |
| 18 | Steel 55 GOST 1050-88 | 210000 | 0.3 |
| 19 | Steel 60 GOST 1050-88 | 204000 | 0.29 |
| 20 | Steel 50G GOST 1050-88 | 216000 | 0.29 |
| 21 | Steel 65G GOST 1050-88 | 215000 | 0.3 |
| 22 | Steel 15Ch GOST 4543-71 | 206000 | 0.29 |
| 23 | Steel 30Ch GOST 4543-71 | 208000 | 0.3 |
| 24 | Steel 35Ch GOST 4543-71 | 214000 | 0.29 |
| 25 | Steel 50Ch GOST 4543-71 | 207000 | 0.29 |
| 26 | Steel 38ChS GOST 4543-71 | 211000 | 0.29 |
| 27 | Steel 35ChM GOST 4543-71 | 213000 | 0.3 |
| 28 | Steel 12Ch18N9T GOST 5949-75 | 195000 | 0.3 |
| 29 | Steel 25L GOST 977-88 | 205000 | 0.3 |
| 30 | Steel 50L GOST 977-88 | 219000 | 0.29 |
| 31 | Grey cast iron 30 GOST1412-85 | 120000 | 0.25 |
| 32 | Grey cast iron 35 GOST1412-85 | 125000 | 0.25 |
| 33 | Nodular cast iron 35 GOST7293-85 | 170000 | 0.21 |
| 34 | Nodular cast iron 60 GOST7293-85 | 180000 | 0.2 |
| 35 | Nodular cast iron 100 GOST7293-85 | 185000 | 0.2 |
| 36 | Malleable cast iron 30-6 | 160000 | 0.25 |

***Source*: compiled by the authors**

According to the geometric constraint, the width value $a$ was calculated depending on the thickness to ensure compliance with the thin plate criterion according to the ratio $a/t \in [20; 100]$, which can be found in [1]. The pressure value was generated on the same principle as for DS1, that is, uniformly in the range $p \in [1000, 10000]$ Pa.

Consequently, a dataset of 3000 unique samples was formed, focussing on practically relevant combinations of real materials and typical steel gauges.

**Finite element modeling.** To obtain the actual deflection values $w$ for the entire range of input parameters in the dataset, a series of calculations using the FEM was conducted. Software was developed combining the Python programming language (using libraries: numpy, pandas, etc.) for input parameter management and the APDL scripting language (Ansys Parametric Design Language) for direct finite element calculations in the Ansys CAE system [22]. The main calculation script contained a main loop iteratively performing calculations and recording results in files for further use in datasets.

The mathematical model, the scheme of which is shown in Fig. 2, represented a thin square plate subjected to a uniformly distributed load.

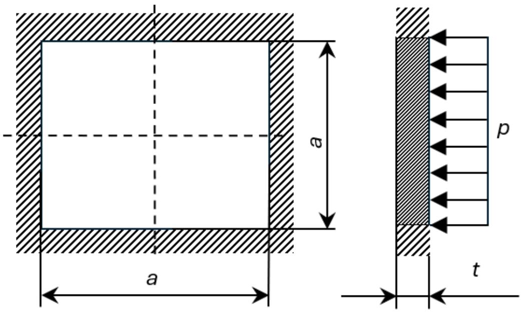



*Fig. 2*. **Calculation scheme of the plate**
***Source*: compiled by the authors**

The material model was defined as elastic and isotropic. Its behaviour is described within the framework of Reissner's geometrically nonlinear theory [1].

The finite element model (see Fig. 3) was built using shell elements.

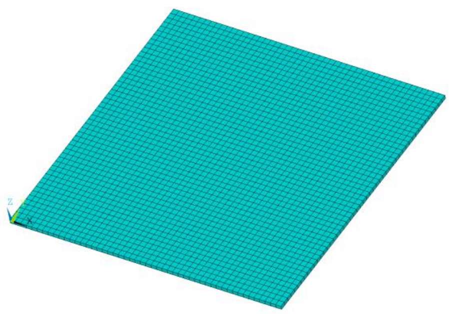


*Fig. 3*. **Geometric model of the plate with FE mesh**
***Source*: compiled by the authors**

A convergence study was previously conducted, resulting in the selection of an optimal mesh density containing 2000 elements, ensuring a balance between accuracy and computational costs.

## TEST MODELS ARCHITECTURE

To ensure objectivity of the results and verify the stability of the proposed data formation strategy, a complex of machine learning algorithm test models (KNN, Random Forest, XGBoost, GPR, and MLP) based on five different mathematical approaches was selected.

This choice allows for the evaluation of the rational dataset's effectiveness for both complex neural networks and classical algorithms sensitive to the density and physical correspondence of input points. All models were trained on the same two datasets (DS1, DS2).

Modelling software was developed in Python using Keras, TensorFlow, scikit-learn, and NumPy. Execution and result viewing were performed using the Jupyter Notebook. Matplotlib was used to visualise the results.

**Random forest model (RF).** The random forest regression algorithm from the scikit-learn library [23] was used as the model implementation. Training was carried out on both datasets (DS1 and DS2). Key hyperparameters were selected using a 5-fold cross-validation on the training set to avoid overfitting. The resulting optimal configuration included 149 decision trees, a maximum tree depth of 23, a minimum of 5 samples required to split an internal node, a minimum of 1 sample per leaf node, and the use of all input features for making split decisions.

**XGBoost model**. The XGBoost algorithm from the scikit-learn library [23] was used to implement the gradient enhancement model. This ensemble algorithm is based on the iterative construction of an ensemble of decision trees. The model parameters were selected to achieve a balance between learning depth and generalisation ability.

The architecture included 500 sequential trees with a maximum depth of up to 6 levels for each tree. Using a relatively low learning rate of 0.05 combined with stochastic feature and sample selection (subsample and colsample_bytree) at a level of 0.8 allowed the model to stably converge to the loss function minimum, avoiding sharp gradient fluctuations.

Regularisation built into the algorithm ensured the stability of deflection prediction even in zones of the feature space with lower data density.

**Gaussian process regression (GPR).** Gaussian process regression was implemented as a probabilistic surrogate model, particularly effective for approximating smooth physical response surfaces.

Since the method is critically sensitive to the scale of the input data, the architecture includes a preliminary feature standardisation stage using the StandardScaler algorithm [23].

The model kernel is constructed as a combination of a constant kernel (initial value 1.0, bounds $10^{-3}$ to $10^{3}$) and a Radial Basis Function (RBF) kernel (initial length scale 1.0, bounds $10^{-2}$ to $10^{2}$), allowing for the modelling of complex nonlinear dependencies between geometric parameters and deflection.

Additionally, a WhiteKernel noise kernel with a noise level of $10^{-5}$ was introduced to account for potential numerical errors in the FE analysis.

Kernel hyperparameter optimisation was performed by maximising logarithmic likelihood with three restarts to exclude falling into local minima.

**K-nearest neighbours (KNN).** The K-Nearest Neighbours model was used as a nonparametric metric approximator. The architecture is based on the principle of estimating the plate response based on the local structure of the dataset.

To correctly calculate Euclidean distances in multidimensional feature space, data normalisation was applied using the StandardScaler algorithm [23]. During the modelling, the optimal number of neighbours was determined as k = 5.

A significant feature of the implementation is the use of a weighting coefficient that is inversely proportional to the distance. This allowed greater weight to be given to samples from the training set that are maximally close to the prediction object in their physical-geometric characteristics, ensuring high local approximation accuracy.

**Multilayer Perceptron (MLP).** The neural network model was based on a multilayer perceptron, the configuration of which was adapted to the specifics of engineering data. The key architecture parameters were established during a global optimisation procedure using a genetic algorithm.

The architecture includes an input layer of 5 neurones, corresponding to the number of input parameters. Four hidden layers of (16, 64, 8, 128) neurones, respectively, and 1 output neurone - deflection *w*. The selected number of neurone levels ensures an optimal balance between computational complexity and the network's generalisation ability, which is critical for working with rational datasets.

This nonmonotonic structure with a “bottleneck” of 8 neurones before expanding to 128 allows the model to form compact high-level representations before final regression. A combined approach was used as the activation function: ReLU for the first hidden layer and Tanh for subsequent layers. This allowed combining the advantages of sparse ReLU activation with the smoothness of Tanh, useful for approximating continuous physical fields. At the output layer, 1 neurone with softplus activation was used to predict the continuous deflection value, ensuring that the prediction of *w* will always be positive, corresponding to the physical meaning of the problem.

To prevent overfitting and ensure model stability, a comprehensive approach was used. The model was compiled using Adam optimiser with an initial learning rate of 0.005 and quality metrics MSE and MAE.

To reduce the chase of potential overfitting associated with such a complex architecture, the regularisation methods with coefficient $\lambda=10^{-5}$ and the dropout with random probability 0.1 disconnection of part of the neurones in the layer at each training step were actively applied.

Training was carried out with a batch size of 32 using the EarlyStopping mechanism, monitoring the error on the validation sample to stop training when the model ceased to improve generalising ability.

## RESEARCH RESULTS

As a result of the developed software operation, two data arrays reflecting different feature space formation strategies were generated.

Fig. 4 shows the matrix of pairwise distributions, allowing for a comprehensive assessment of the structure of the formed datasets.

Primary statistical analysis revealed significant structural differences between them. For the synthetic dataset DS1, a uniform distribution of all quantities is observed, confirming the operation of the correctness of the Sobol algorithm. On the diagonal, kernel density estimation (KDE) plots are shown, where for the rationally formed set DS2, the multimodality of the distributions for the parameters *t*, *E* and *ν* is clearly traced, corresponding to the discrete gauge and specific steel grades.

Projections of points below the diagonal visualise the imposed geometric and material science constraints, particularly the clustering of material properties and the wedge-shaped region of permissible geometric parameters *a* – *t*. At the intersection of parameters *a* and *t*, a wedge-shaped region of permissible values is clearly traced, bounded by lines $a = 20t$ and $a = 100t$, corresponding to the limitations imposed on the ratio for thin plates described in the previous section, which is a direct consequence of integrating physical knowledge a priori into the dataset.

Analysis of correlations (Fig. 5) showed that in the synthetic dataset DS1, the Pearson correlation coefficients between all pairs of input parameters are close to zero, indicating their statistical independence. In the DS2 dataset, a strong positive correlation ($r$ = 0.77) was found between the width of the plate *a* and thickness *t*.


ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

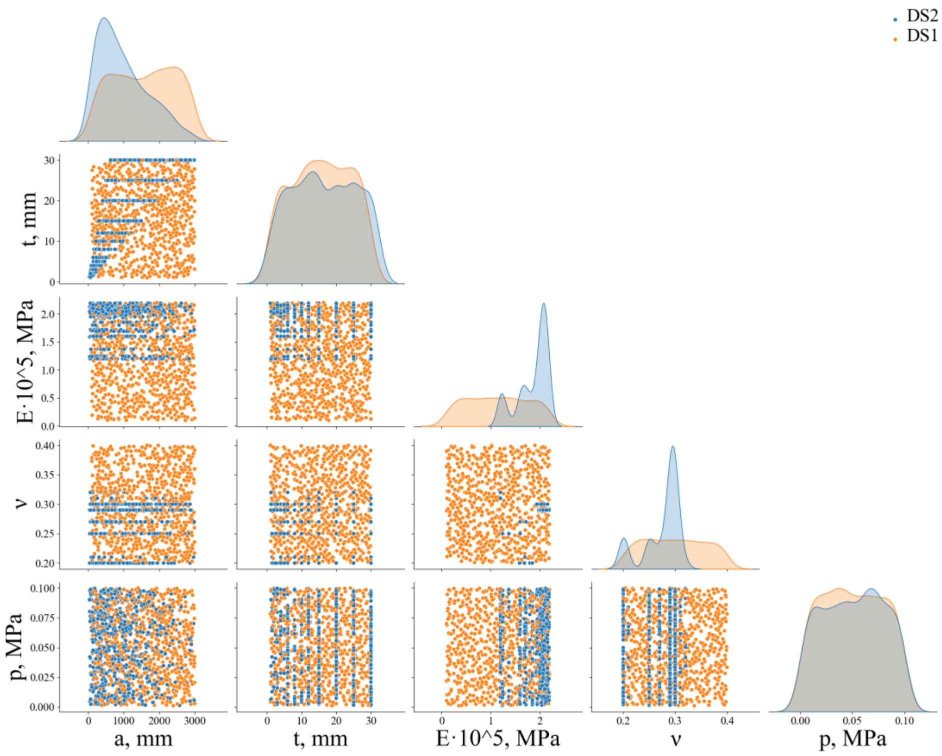
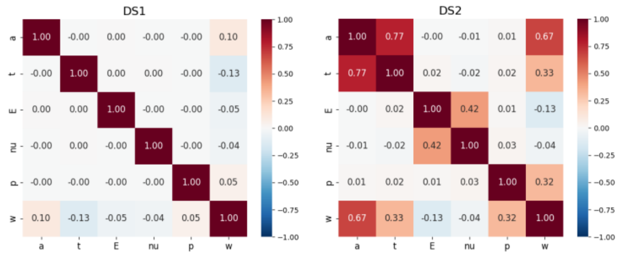



*Fig. 4.* **Scatter plot matrix 5x5 of pairwise distributions for input parameters. Colours indicate datasets: DS1 (orange) and DS2 (blue)**
***Source*: compiled by the authors**




a b

*Fig. 5.* **Comparative analysis of correlation matrices for visualising dataset structural differences:**
**a – DS1; b – DS2. Red colour – positive correlation, blue – negative**
***Source*: compiled by the authors**

These correlations are a consequence of the geometric constraint *a/t* ∈ [20, 100]. Thus, DS1 represents an abstract, unconnected parameter “hypercube”, while DS2 is a structured, correlated, and clustered subspace reflecting imposed physical constraints and engineering practice realities.

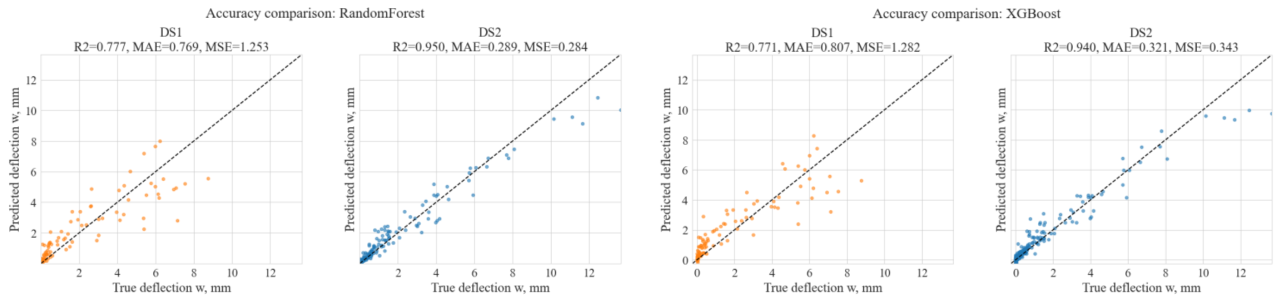
**Evaluation of the accuracy and efficiency of surrogate models.** After conducting 3000 numerical FEM simulations for each strategy, the two datasets described above (DS1 and DS2) were supplemented with the corresponding deflection values in the centre of the plate *w*. Thus, two complete feature spaces suitable for further machine learning and surrogate model development were formed.

Model training was conducted under identical conditions for both datasets. The general behaviour of the dynamics of the loss function demonstrated that test models on synthetic data (DS1) quickly reach a certain error level, after which they plateau with moderate accuracy (e.g., for MLP MAE ≈ 0.2-0.25 mm). On the contrary, training with rationally formed data (DS2) proceeded more stably, allowing a significantly lower residual error level (e.g., for the MLP model MAE ≈ 0.08 mm).

This confirms the hypothesis that structured data accounting for physical constraints allow models to find a more accurate and stable global minimum. Scatter plots (Fig. 6, Fig. 7 and Fig. 8) of predicted values relative to actual values visualise this difference.

It should be noted that for models trained on DS1, significant point scatter is observed, especially in the region of small deflections. Models based on DS2 demonstrate a high grouping density of points along the ideal diagonal, indicating high prediction reliability throughout the entire working range.

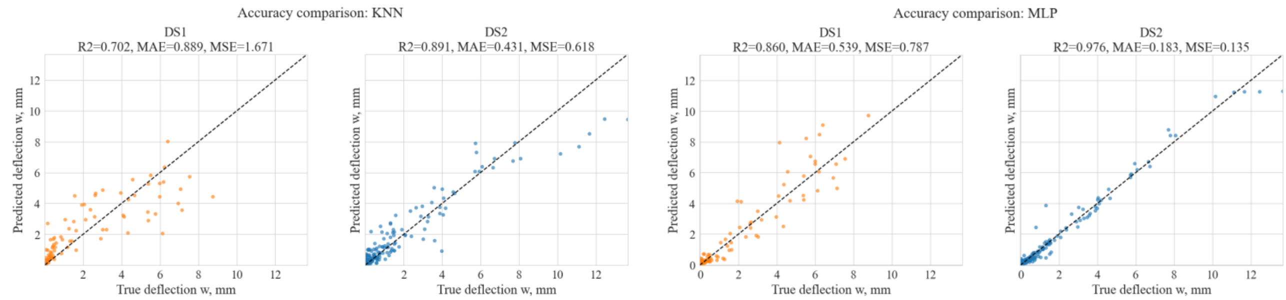
The best result on the rational data set was shown by the neural network $R^2 = 0.977$ and the GPR model $R^2 = 0.99$, surpassing the ensemble methods with $R^2 = 0.950$ for RF and $R^2 = 0.940$ for XGBoost.




a b

*Fig. 6.* **Comparison of the predicted and actual deflection values:**
**a – for the RF model; b – for the XGBoost model**

***Source*: compiled by the authors**




a b

*Fig. 7.* **Comparison of the predicted and actual deflection values:**
**a – for the KNN model; b – for the MLP model**

***Source*: compiled by the authors**

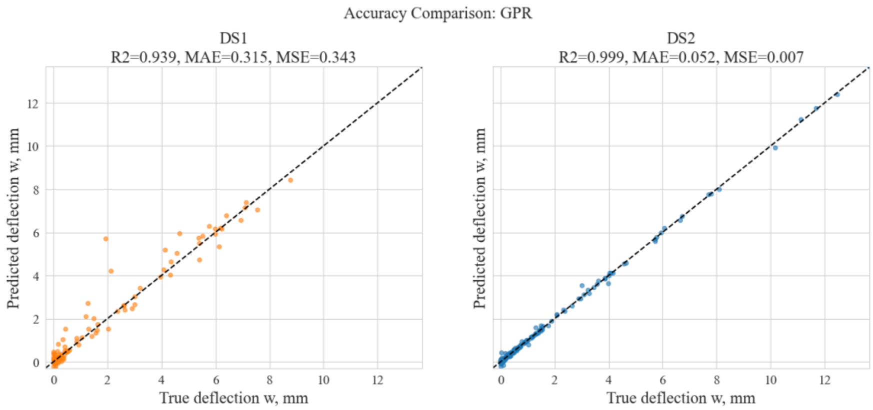



*Fig. 8.* **Comparison of predicted and actual deflection values for the GPR model**

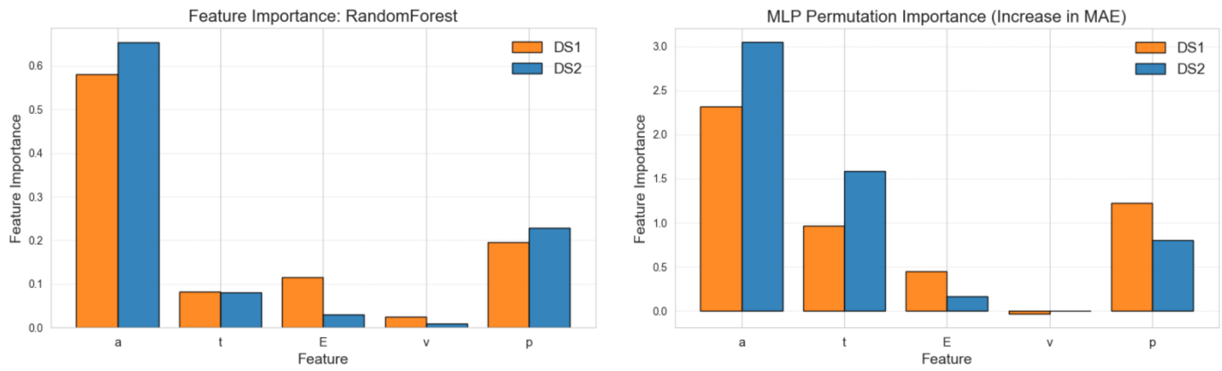
***Source*: compiled by the authors**

**Model interpretation and feature importance analysis.** To verify the physical adequacy of the models, feature importance analysis was performed for both datasets. Results for two characteristic model types, RF (as a representative of ensemble models) and MLP, are shown below in Fig. 9.

For the ensemble model (Fig. 9a), the dominant factors were determined to be the plate width *a* and pressure *p*. At the same time, the material parameters *E* and *ν* have less importance.

The influence of thickness *t* was assessed as insignificant (weight < 0.1), which contradicts the physical theory of plate bending, where deflection is inversely proportional to the thickness cube ($w \approx 1/t^3$).This indicates that models based on ensemble methods, despite high $R^2$, are unable to fully reproduce the physical nature of the phenomenon.

On the contrary, the feature importance analysis for the neural network (Fig. 9b), performed using the permutation importance method, showed a fundamentally different picture. The model assigned high significance to the thickness parameter *t* (about 1.5 for DS2), which correlates with analytical dependencies. This suggests that the deep architecture of the neural network, trained on structured data, is capable of better approximating complex nonlinear dependencies than decision tree-based algorithms.




a b

*Fig. 9.* **Feature importance:**
**a – for the RF model; b – for the neural network**
***Source*: compiled by the authors**

A general quantitative comparison of results on the test sample for both datasets and all model types (MLP, RF, XGBoost, KNN, GPR) is given in Table 3 and Fig. 10 and Fig.11.

*Table 3.* **Comparison of model quality in the test sample**

| Model | Dataset | MSE | MAE | $R^2$ |
|---|---|---|---|---|
| RF | DS1 | 1.25270 | 0.76890 | 0.77655 |
| **RF** | **DS2** | **0.28368** | **0.28865** | **0.95003** |
| XGBoost | DS1 | 1.28217 | 0.80679 | 0.77129 |
| **XGBoost** | **DS2** | **0.34305** | **0.32110** | **0.93958** |
| KNN | DS1 | 1.67145 | 0.88938 | 0.70186 |
| **KNN** | **DS2** | **0.61813** | **0.43092** | **0.89112** |
| GPR | DS1 | 0.34268 | 0.31525 | 0.93887 |
| **GPR** | **DS2** | **0.00728** | **0.05166** | **0.99872** |
| MLP | DS1 | 0.78749 | 0.53904 | 0.85953 |
| **MLP** | **DS2** | **0.13532** | **0.18341** | **0.97617** |

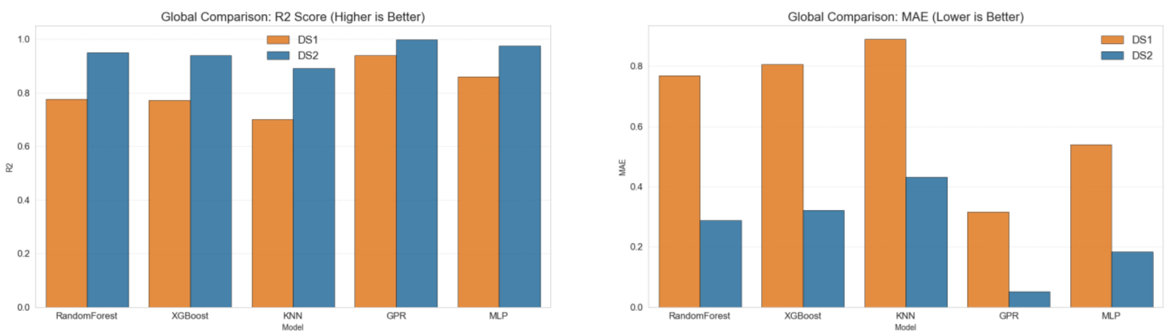
***Source*: compiled by the authors**

## DISCUSSION OF RESULTS

The results obtained during the study allow for a reevaluation of the role of data sampling strategy in surrogate modelling tasks for mechanical systems. The key factor that determined the significant difference in model efficiency was the fundamental structural difference in the training samples.

As shown by the analysis of distributions and correlations, the synthetic approach (DS1) forms an abstract parameter hypercube, which is mathematically correct from the perspective of experimental design theory but physically redundant. On the contrary, the rational approach (DS2), based on physical, geometric, and material science constraints, creates a sample reflecting real engineering practice, where parameters are discrete and interdependent.

In contrast, incorporating engineering knowledge into dataset DS2 effectively acts as a form of “hard data regularisation”. This deliberately introduces a certain bias towards real design solutions, which, as confirmed by the results ($R^2$ = 0.9762), significantly stabilises the variance and ensures the high robustness of the model within its operational range. Such search-space constraints allow computational resources to be focused on approximating real physical effects rather than attempting to interpolate anomalous structural states.




a b

*Fig. 10.* **Comparison of accuracy for both datasets for each test model:**
**a – $R^2$; b – MAE**
***Source*: compiled by the authors**

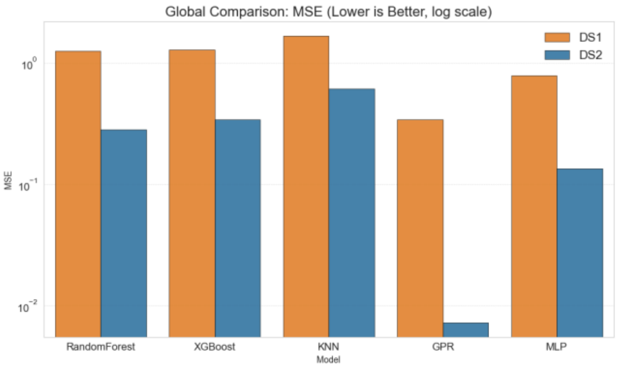



*Fig. 11*. **Comparison of MSE accuracy for both datasets for each test model**
***Source*: compiled by the authors**

An important result is the empirical confirmation of the advantage of deep neural networks over ensemble methods in solid mechanics tasks.

Comparative analysis of feature importance demonstrated that ensemble algorithms, despite their popularity, are unable to fully reproduce the strong nonlinear (cubic) dependence of deflection on plate thickness, underestimating the weight of this parameter. This limitation is inherent to the mathematical nature of decision tree-based models, which function as piecewise-constant approximators. Consequently, they cannot naturally form smooth high-order polynomial response surfaces (such as the inverse cubic dependence $w \approx 1/t3$) without an excessive number of splits. In contrast, the MLP architecture with non-linear activation functions successfully approximated this physical regularity.

The proposed rational approach directly addresses the problem of generating physically impossible parameter combinations (non-existent materials) discussed in the literature review. Instead of expending the approximator's computational resources on learning the behaviour of non-existent compounds, the model focusses on a narrow but practically significant range of structural steel properties. This shows that the quality and physical adequacy of the training data is a factor of comparable importance to the model architecture itself.

At the same time, it is worth noting the limitations of the results obtained. Detailed error analysis revealed an increase in relative prediction error for small deflections (<0.1 mm). This is explained by the fact that the MSE loss function optimises the absolute error, which for small values contributes insignificantly to the total gradient. This limits the application of the developed model for precision micromechanics tasks but leaves it highly effective for calculating engineering structures and elements of machines and mechanisms where significant deformations are critical.

Further development of the research is seen in extending the proposed sampling approach to tasks considering plastic deformations and geometric nonlinearity, where the effect of "physical" parameter space filtration may be even more pronounced due to the higher sensitivity of such systems to initial conditions.

## CONCLUSIONS

The study conducted a comprehensive investigation into the impact of training data generation approaches on the efficiency of surrogate modelling of the stress-strain state. The square plate problem served as a valid benchmark to evaluate the proposed tools against classical methods. Based on the results obtained, the following conclusions can be drawn.

A system of geometric and physical-mechanical constraints was formalised, allowing for the exclusion of non-physical states from the parameter space at the experimental design stage. On this basis, algorithmic and software tools were developed for automated data generation using two strategies: synthetic (quasirandom) and rational, integrating standard metal assortment parameters and real properties of structural steels.

The effectiveness of preventive integration of engineering knowledge was proven. This approach allows the search space to be narrowed to a physically plausible subset, eliminating the problem of training models on realistic parameter combinations. Analysis of the results of numerical modelling revealed that in the DS1 sample, more than 52 % of generated samples fell outside the applicability limits of small deflection theory. On the developed rational strategy DS2 ensured 98% relevant samples, allowing radical optimisation of computational costs even at the dataset formation stage.

A radical increase in predictive accuracy was established for all investigated surrogate models when transitioning to the rational formed dataset. Specifically, for the MLP neural network, the coefficient of determination $R^2$ increased from 0.8595 (for DS1) to 0.9762 (for DS2), and the mean absolute error decreased almost three times – from 0.5390 to 0.1834. A similar trend is observed for other algorithms, where the precision of the KNN model increased from 0.7019 to 0.8911, and RF demonstrated the growth of $R^2$ from 0.7766 to

 ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

0.9500. The highest efficiency was shown by Gaussian process regression (GPR), reaching an indicator of $R^2 = 0.9987$ in the rational dataset.

The priority of data quality over architecture complexity was substantiated. The results obtained serve as empirical confirmation that the physical consistency of the training sample is a dominant factor in the construction of reliable intelligent systems, surpassing the complexity of the model architecture itself in importance.

Comparative analysis of feature importance showed the advantage of deep neural networks over ensemble methods. Specifically, MLP successfully identified the nonlinear influence of plate thickness on its stiffness, while the RF model demonstrated a tendency to ignore this fundamental parameter.

The practical value of the study was determined. The proposed "data-centric" approach to creating surrogate models can be recommended for developing predictive maintenance and rapid diagnostic systems. While demonstrated on a fundamental plate problem, this proposed sampling procedure provides a proven basis for future application to complex non-linear tasks, such as the analysis of shells with dents, where minimizing computational costs is critical. This ensures a combination of high accuracy and real-time prediction speed with complete physical adequacy of the results.

## REFERENCES

1. Ozenda, O., Virga, E. G. "On the Kirchhoff-Love hypothesis (revised and vindicated)". *Journal of Elasticity*. 2021; 143 (2): 359–384, https://www.scopus.com/pages/publications/85100816179. DOI: https://doi.org/10.1007/s10659-021-09819-7.

2. Ventsel, E., Krauthammer, T. "Thin plates and shells: Theory, analysis, and applications". *Marcel Dekker.* 2001. DOI: https://doi.org/10.1201/9780203908723.

3. Zienkiewicz, O. C., Taylor, R. L. & Govindjee, S. "The finite element method: Its basis and fundamentals". *8th ed. Butterworth-Heinemann.* 2024.

4. Bathe, K. J. "Finite element procedures". *2nd ed. Klaus-Jurgen Bathe*. 2014.

5. Sajjadinia, S. S., Carpentieri, B. & Holzapfel, G. A. "A pointwise evaluation metric to visualize errors in machine learning surrogate models". *Frontiers in Artificial Intelligence and Applications.* 2021; 345: 26–34, https://www.scopus.com/pages/publications/85123633675. DOI: https://doi.org/10.3233/FAIA210386.

6. Sajjadinia, S. S., Carpentieri, B., Shriram, D. & Holzapfel, G. A. "Multi-fidelity surrogate modeling through hybrid machine learning for biomechanical and finite element analysis of soft tissues". *Computers in Biology and Medicine.* 2022; 148: 105699, https://www.scopus.com/pages/publications/85132910358. DOI: https://doi.org/10.1016/j.compbiomed.2022.105699.

7. Babudzhan, R. A., Vodka, O. O. & Shapovalova, M. I. "Application of computational intelligence methods for the heterogeneous material stress state evaluation". *Herald of Advanced Information Technology*. 2022; 5 (3): 198–209, https://www.scopus.com/pages/publications/105024884770. DOI: https://doi.org/10.15276/hait.05.2022.15.

8. Hastie, T., Tibshirani, R. & Friedman, J. "The elements of statistical learning: Data mining, inference, and prediction". *2nd ed. Springer*. 2009.

9. Choporova, O. V. & Lisniak, A. O. "The use of a genetic algorithm for optimizing the parameters of the neural network in predicting the stress-strain state of a squared plate". *Applied Questions of Mathematical Modelling.* 2020; 3 (2.1): 164–174. DOI: https://doi.org/10.32782/KNTU2618-0340/2020.3.2-1.27.

10. Azimi, M., Eslamlou, A. D. & Pekcan, G. "Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review". *Sensors*. 2020; 20 (10): 2778, https://www.scopus.com/pages/publications/85084786387. DOI: https://doi.org/10.3390/s20102778.

11. 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, https://www.scopus.com/pages/publications/85057501376. DOI: https://doi.org/10.1016/j.jcp.2018.10.045.

12. Krishnapriyan, A. S., Gholami, A., Zhe, S., Kirby, R. M. & Mahoney, M. W. "Characterizing possible failure modes in physics-informed neural networks". *Advances in Neural Information Processing Systems.* 2021; 34: 26548–26560. DOI: https://doi.org/10.48550/arXiv.2109.01050.

13. Onatskyi, R. & Misiura, S. "Development of a methodology for rapid strength assessment of dented pressure vessels using AI: dataset formation for plates under pressure". *Bulletin of the National Technical University "KhPI". Series: Dynamics and Strength of Machines.* 2025; 1: 60–68. DOI: https://doi.org/10.20998/2078-9130.2025.1.331577.

14. Metropolis, N. & Ulam, S. "The Monte Carlo method". *Journal of the American Statistical Association.* 1949; 44 (247): 335–341. DOI: https://doi.org/10.1080/01621459.1949.10483310.

15. Sobol, I. M. "On the distribution of points in a cube and the approximate evaluation of integrals". *USSR Computational Mathematics and Mathematical Physics.* 1967; 7 (4): 86–112. DOI: https://doi.org/10.1016/0041-5553(67)90144-9.

16. Saltelli, A., Annoni, P., Azzini, I., et al. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". *Computer Physics Communications.* 2010; 181 (2): 259–270, https://www.scopus.com/pages/publications/70450224663.
DOI: https://doi.org/10.1016/j.cpc.2009.09.018.

17. Thaler, K., Schneider, R. & Bucher, C. "Training data selection for machine learning-enhanced Monte Carlo simulations in structural dynamics". *Applied Sciences*. 2022; 12 (2): 581, https://www.scopus.com/pages/publications/85122267929. DOI: https://doi.org/10.3390/app12020581.

18. Borisut, P. & Nuchitprasittichai, A. "Adaptive Latin Hypercube Sampling for a Surrogate-Based Optimization with Artificial Neural Network". *Processes*. 2023; 11 (11): 3232, https://www.scopus.com/pages/publications/85178349811. DOI: https://doi.org/10.3390/pr11113232.

19. Dodwell, T. J., Girolami, M. & Rogers, T. "A data-centric approach to generative modelling for 3D-printed steel". *Proceedings of the Royal Society A.* 2021; 2255: 20210444, https://www.scopus.com/pages/publications/85122328270. DOI: https://doi.org/ 10.1098/rspa.2021.0444.

20. Weerasinghe, G., Kannan, R. & Banda, C. "Making global sensitivity analysis feasible using neural network surrogates". *Data-Centric Engineering.* 2025; 7: e1, https://www.scopus.com/pages/publications/105024554800. DOI: https://doi.org/10.1017/dce.2025.10029.

21. Nawale, P. "Materials Dataset". *Kaggle.* 2023. – Available from: https://www.kaggle.com/datasets/purushottamnawale/materials. – [Accessed: Jan 2026].

22. "ANSYS Mechanical APDL Programmer's Reference. Release 25.R1". *ANSYS, Inc*. 2025. – Available from: https://www.ansys.com. – [Accessed: Dec 2025].

23. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. "Scikit-learn: Machine learning in Python". *Journal of Machine Learning Research.* 2011; 12: 2825–2830.

**Conflicts of Interest:** The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 08.01.2026
Received after revision 13.03.2026
Accepted 19.03.2026

**DOI: https://doi.org/10.15276/aait.09.2026.11**
**УДК 004.85:539.3**

# Раціональне формування навчальних вибірок для сурогатного моделювання напружено-деформованого стану пластинчастих елементів

**Онацький Роман Леонідович**[1)]
ORCID: https://orcid.org/0009-0004-6305-8442; roman.onatskyi@infiz.khpi.edu.ua
**Місюра Сергій Юрійович**[1)]
ORCID: https://orcid.org/0000-0002-5048-1610; serhii.misiura@khpi.edu.ua. Scopus Author ID: 56416492100
[1)] Національний технічний університет «Харківський політехнічний інститут», вул. Кирпичова, 2. Харків, 61002, Україна

## АНОТАЦІЯ

**Актуальність.** Сурогатні моделі на основі машинного навчання є перспективною альтернативою обчислювально витратному методу скінченних елементів для прогнозування напружено-деформованого стану конструктивних елементів. Однак їхня ефективність критично залежить від якості навчальних даних, тоді як більшість існуючих підходів спирається на абстрактні стохастичні методи генерації вибірок, що ігнорують фізичну природу інженерних параметрів. **Мета і завдання.** Метою дослідження є порівняльний аналіз ефективності стохастичних та фізично-орієнтованих стратегій формування навчальних даних для сурогатного моделювання напружено-деформованого стану пластинчатих елементів. Фокус роботи зміщено з ускладнення архітектури моделей на розробку підходу до раціонального формування датасетів, що інтегрують апріорні інженерні знання. **Методи.** На прикладі задачі сурогатного моделювання прогину тонкої жорстко закріпленої пластини порівнюються дві фундаментально різні стратегії формування вибірок. Класична, на основі квазі-випадкових послідовностей Соболя, та запропонована раціональна. Остання базується на реальних емпіричних властивостях конструкційних сталей та стандартизовані товщини листового прокату. Для оцінки ефективності та кількісного порівняння впливу структури даних на точність прогнозування розроблено процедуру інтелектуального моделювання. Як тестові моделі для задачі регресії було використано наступні алгоритми машинного навчання: метод K-найближчих сусідів, випадкового лісу, XGBoost, регресії гауссового процесу та багатошарового перцептрону. **Результати.** Встановлено, що перехід до раціональної стратегії забезпечує стабільне зниження середньоквадратичної похибки для всіх досліджених моделей у три з половиною – чотири рази порівняно з базовою стратегією. Використання раціонального датасету підвищило коефіцієнт детермінації на двадцять сім відсотків для методу k-найближчих сусідів, на двадцять три відсотки для випадкового лісу та на двадцять один відсоток для XGBoost. Найвищу точність показали моделі регресії на основі гауссових процесів (із коефіцієнтом детермінації дев'яносто дев'ять відсотків) та багатошаровий перцептрон (із коефіцієнтом детермінації дев'яносто вісім відсотків). Доведено, що низька ефективність традиційних вибірок обумовлена тим, що понад п'ятдесят відсотків зразків містять нефізичні комбінації геометричних розмірів та властивостей матеріалів, які не відповідають жодним реальним інженерним стандартам. Ці нереалістичні комбінації параметрів створюють «інформаційний шум» під час навчання. **Висновки.** Запропонований підхід раціонального формування вибірок слугує основою для створення робастних ШІ-інструментів експрес-діагностики інженерних конструкцій, забезпечуючи високу узагальнюючу здатність моделей при меншому обсязі вхідних даних.

**Ключові слова:** напружено-деформований стан; сурогатне моделювання; регресійний аналіз; машинне навчання; раціональна стратегія формування даних; датасет; швидка діагностика інженерних конструкцій

## ABOUT THE AUTHORS

**Roman L. Onatskyi** - PhD student Department of Mathematical Modelling and Intelligent Computing in Engineering. National Technical University “Kharkiv Polytechnic Institute”, 2, Kyrpychova Str. Kharkiv, 61002, Ukraine
ORCID: https://orcid.org/0009-0004-6305-8442; roman.onatskyi@infiz.khpi.edu.ua
***Research field*:** computational mechanics, stress-strain state prediction, residual life assessment of shell structures with defects, machine learning in solid mechanics, surrogate modeling, supervised learning, data-driven engineering

**Онацький Роман Леонідович** - аспірант кафедри Математичного моделювання та інтелектуальних обчислень в інженерії. Національний технічний університет «Харківський політехнічний інститут», вул. Кирпичова, 2. Харків, 61002, Україна

**Seghii Yu. Misiura** - Candidate of Engineering Sciences, Senior researcher, Associate Professor, Department of Mathematical Modelling and Intelligent Computing in Engineering. National Technical University “Kharkiv Polytechnic Institute”, 2, Kyrpychova Str. Kharkiv, 61002, Ukraine
ORCID: https://orcid.org/0000-0002-5048-1610; serhii.misiura@khpi.edu.ua. Scopus Author ID: 56416492100
***Research field:*** strength, vibrations and optimal design of structures, computational mechanics, stress-strain state prediction, residual life assessment of shell structures with defects, machine learning in solid mechanics, surrogate modeling, supervised learning, data-driven engineering

**Місюра Сергій Юрійович** - кандидат технічних наук, доцент, старший дослідник, доцент кафедри Математичного моделювання та інтелектуальних обчислень в інженерії. Національний технічний університет «Харківський політехнічний інститут», вул. Кирпичова, 2. Харків, 61002, Україна